OCT 17 1922

The Journal of Educational Psychology

Devoted Primarily to the Scientific Study of Problems of Learning and Teaching.

BOARD OF EDITORS:

HAROLD ORDWAY RUGG, *Chairman.*
Lincoln School of Teachers College.
Teachers College, Columbia University.

JAMES CARLETON BELL,
Brooklyn Training School for Teachers.

FRANK NUGENT FREEMAN,
University of Chicago.

ARTHUR IRVING GATES,
Teachers College, Columbia University.

VIVIAN ALLEN CHARLES HENMON,
University of Wisconsin.

RUDOLF PINTNER,
Teachers College, Columbia University.

BEARDSLEY RUML,
Carnegie Corporation, New York City.

LEWIS MADISON TERMAN,
Leland Stanford University.

EDWARD LEE THORNDIKE,
Teachers College, Columbia University.

LAURA ZIRBES, *Assistant Editor.*
Lincoln School of Teachers College.

VOL. XIII.
No. 7.

OCTOBER, 1922

$4.00 a year.
60c. per copy.

CONTENTS

QUOTIENTS I, E, AND A. *Rena Stebbins and L. A. Pechstein.* 385

THE TEACHING OF EDUCATIONAL PSYCHOLOGY IN THE UNITED STATES. *H. H. Remmers and F. B. Knight* 399

A COMPARISON OF MENTAL AGE SCORES OBTAINED BY PERFORMANCE TESTS AND THE STANFORD REVISION OF THE BINET-SIMON SCALE. *Buford Johnson and Louise Schriefer* . . . 408

CORRELATIONS OF FOUR INTELLIGENCE TESTS WITH GRADES. *A. M. Jordan* 419

LANGUAGE ERROR TESTS. *G. M. Wilson* 430

THE RELATION OF RHYTHM TO THE HANDWRITING MOVEMENT. *Paul V. West.* 438

NEW PUBLICATIONS IN EDUCATIONAL PSYCHOLOGY AND RELATED FIELDS OF EDUCATION. 445

Published Monthly Except June to August by
WARWICK and YORK, Inc.,
York, Pa. *Baltimore, Md.*

Entered as Second Class matter November 15, 1921, at the Post Office at York, Pennsylvania, under the Act of March 3, 1879.

The Journal of Educational Psychology

is published monthly, by Warwick & York, Inc., Baltimore, Md., and York, Pa. The issues for 1922 make Volume XIII. Title page and index are bound in December number.

Manuscripts for publication, books or other materials for review, and news items should be addressed to Harold Ordway Rugg, Chairman of the Editorial Board, 59 Edgecliffe Terrace, Yonkers, New York.

The price of the Journal is $4.00 a year, payable in advance. Foreign postage is 30 cents extra. Single issues are 60 cents each, and less than a full year's subscription is charged at single issue rate for every month ordered.

Subscribers should notify the publishers of change in address at least four weeks before publication of the issue with which change is to take effect. No claim for non-receipt of an issue will be entertained unless made within two weeks after receipt of the next succeeding issue.

Warwick & York, Inc.,
Baltimore, Md.

TEACHERS' HEADQUARTERS

European
Plan

Entirely
Fire-Proof

Centrally
Located

EDWARD DAVIS
Manager

HOTEL RENNERT
BALTIMORE, MARYLAND

RECENT PUBLICATIONS

Brief Introductory Psychology for Teachers.

By Edward K. Strong, Jr.

xii, 241 *pages.* $1.60. *Postage* 10¢.

The Output of Professional Schools for Teachers.

By Charles E. Benson.

x, 88 *pages.* $1.50. *Postage* 10¢.

Textbook Selection.

By R. H. Frazen and F. B. Knight.
Introduction by Ernest Horn.

94 *pages.* $1.20. *Postage* 6¢.

Warwick & York, Inc., Publishers, Baltimore.

THE JOURNAL OF
EDUCATIONAL PSYCHOLOGY

Vol. XIII October, 1922 No. 7

QUOTIENTS I, E, AND A

RENA STEBBINS

AND

L. A. PECHSTEIN

City Normal School, Rochester, N. Y., Department of Psychology and Education, University of Rochester

When an examiner has finished giving a battery of educational tests in a modern school, he has two groups of data to offer the principal —first, a record of grade norms in comparison with standard norms, showing what grades have central tendencies above, equal to, or below standard; second, individual class record sheets for each test, showing what score each child made on each test. And as the examiner gives this material to the teacher, to the supervisor, or to the principal, he is struck with the unfairness of it all—its unfairness to the teacher, because the records show only how her grade compares in measurable educational abilities with other grades in the country, without taking into account the pupil-material with which she worked; its unfairness to the pupil, because it gives high rank and praise to those above the norm and a sense of failure and blame to those below the norm, without taking into account the native capacity or intelligence of these children.

Those interested in educational testing have long realized that the valuable and fair testing program will be one in which the results of educational tests are compared with the results of mental tests; in which we measure both teacher and pupil not by what the child has learned, but by what he has learned in relation to what he is capable of learning. The need for a combination of educational and mental tests has long been felt. The method has been the chief difficulty. Several different lines of approach have been taken in various parts of the country. In Rochester, we have chosen the method which follows because of its simplicity, its fairness, and its possibilities for interpretation.

The School Situation

Four hundred children in four grades in the Practice Department of the City Normal School, Rochester, N. Y., were chosen for our field of testing. Because of other testing projects in the first three grades, we limited our work to Grades IV through VII. Each year of school work is divided in two, and promotions are half-yearly. The first half-year in a grade is designated by *B*, the second half-year by *A*. Not only are there *B* and *A* grades for each year, but in every grade there are two parallel classes, one composed of superior children, one of inferior children, classified on the basis of teachers' judgments. Thus there are in the fourth grade:

IV B_1—First half of Grade IV, brighter pupils
IV B_2—First half of Grade IV, slower pupils
IV A_1—Second half of Grade IV, brighter pupils
IV A_2—Second half of Grade IV, slower pupils

In the four grades studied there were in all, then, 16 classes, taught by 14 well-trained teachers.

Selection and Administration of Tests

In selecting tests for our educational survey, no attempt was made to choose new tests for the sake of evaluating them. Only well-known, widely used tests were selected, tests whose value had already been clearly demonstrated elsewhere. The tests chosen were:

Mental
National Intelligence Test A and B, Form I.
Educational
Reading Thorndike-McCall
Arithmetic Fundamentals { Woody Addition Scale A; Woody Subtraction Scale A; Woody Multiplication Scale A; Woody Division Scale A
Problems Buckingham Problem Scale
Spelling { Ayres' (in lists); Monroe Timed—Sentence Spelling Test.

The group mental tests were given by Miss Leila Martin, Director of the Child Study Bureau of Rochester, under excellent and uniform conditions. The educational tests were all given by one trained examiner under uniform testing conditions, and all corrections, tabulations and statistical work were done either by this examiner or

under her personal supervision. Because of this great care in the administration of the tests, we feel that mechanical error has been reduced to a minimum and that conditions warrant a close comparison between grades.

The Scores

When the tests were given, we had scores in two group intelligence tests and in seven educational tests for each child. The typical record for each child follows:

A. J.—Age 10 Years 3 Months	Score
National Intelligence Scale A	94.0
National Intelligence Scale B	96.0
Thorndike-McCall Reading Test	46.0
Woody Addition Scale A	28.0
Woody Subtraction Scale A	23.0
Woody Multiplication Scale A	27.0
Woody Division Scale A	24.0
Buckingham Problem Scale	6.46
Ayres' Spelling Scale	77.0
Monroe Timed—Sentence Spelling Test	42.0

One glance at this record makes the difficulty apparent. When these scores are given to the teacher, they are valuable only insofar as she may compare them with scores of other children in her grade. This comparison she had already made, however, long before we gave the tests. A more concrete comparison she cannot make. If these scores should ever be given to a parent, he would get almost no information from them, except the judgment that he, who used to be "good in reading" has a child who "stands 46." Even the examiner himself cannot interpret or compare results based upon such varying criteria without further work. Clearly before we can study the relation of this child's educational achievement to his mental capacity, we must reduce these unrelated scores to some comparable unit. The method of composite scores, and the method of composite ranks seem to yield results less capable of interpretation to the average teacher and parent than the method of the educational age.

Mental and Educational Ages

In the field of intelligence tests, all efforts to substitute scores, percentiles, intelligence indices, etc., for Binet's and Terman's respective contributions of the mental age (MA) and the intelligence quo-

tient (IQ) have failed. The layman understands MA and he wishes the results of the intelligence test stated in terms which he can comprehend. In the field of educational measures, if we could give both the teacher and parent in the case of the child mentioned above, instead of a group of scores, a statement like this: "The child reads as well as the average child of 11 years 6 months; in addition, he does as well as the average child of 11 years 10 months; in subtraction, 11 years 6 months; in multiplication, 12 years; in division, 11 years 10 months; in problems, 12 years; in list spelling, 12 years 4 months," we would make test results immeasurably more valuable to teacher, pupil, and parent. This is what we have attempted to do with the scores obtained in our study.

The transforming of national intelligence scores into mental ages was simple. Using the age norms given by the National Research Council, we interpolated for months, forming an age-score table as follows:

PROVISIONAL AGE STANDARDS
National Intelligence Test Form A-1

Years	Months	Score	Years	Months	Score	Years	Months	Score	Years	Months	Score
8	0	65.0	9	0	78.0	10	0	91	11	0	103.0
8	1	66.1	9	1	79.1	10	1	92	11	1	103.8
8	2	67.2	9	2	80.2	10	2	93	11	2	104.6
8	3	68.3	9	3	81.3	10	3	94	11	3	105.5
etc.			etc.			etc.			etc.		

From these tables we read the mental ages which corresponded to the scores on Scales A and B, and averaged them to find the child's mental age. While no one experienced in the inaccuracies of group testing at its best would care to state that this is the child's true mental age, at least it is the mental age which corresponds to his scores obtained on group tests given under the best possible conditions.

Our next task was to transform educational scores into what we have called an educational age. If we say that a child has a mental age of 12 years when he does as well on a mental test as the average child of that age, then we may say that a child has an "educational age" of 12 years when he does as well on educational tests as the average child of that age. It is clear that in order to transpose scores into educational ages the same method of standardization is necessary that is used in finding age norms for mental tests. Large unselected

groups of children at each age, taken regardless of the grade in which they are located, should be given the test, and age standards found from the median or mean performance of these children. This has been done in the case of the Thorndike-McCall Reading Scales. In a clear discussion of scale construction Dr. McCall[1] points out how he procured the educational age norms for his 10 reading scales. The examiner, by using the table for reading age standards sent with the manual of directions, can find directly and accurately the "reading age" for each child. This has been done with all our reading scores.

Method of Deriving Educational Ages

For the other tests in our study, only grade norms were furnished by the test originators. Our problem was to find some method of using these grade norms given with the tests to find educational ages. We realize that the method we have used has its defects; that it is at best only a makeshift; that it is as accurate as it could now be made and, until test constructors furnish us with age norms, it is the best method for interpreting scores.

Let us take for an example the Buckingham problem test. Dr. Buckingham has found the following grade medians for the end of the term, based on large numbers of pupils tested:

End of	Score
III*B*	3.66
III*A*	3.84
IV*B*	4.36
IV*A*	4.74
etc.	

It is generally accepted by educators, and the 1918 school census reports support the facts, that the standard median ages for the above grades are:

End of	Years	Months
III*B*	9	
III*A*	9	6
IV*B*	10	
IV*A*	10	6

We have said, since the median performance in the problem scale for III*B* grade is 3.66, and since the median age for III*B* grade is 9

[1] McCall, William A.: Uniform Method of Scale Construction, *Teachers College Record*, Jan., 1921.

years, we shall let 9 years be the "problem age" for all children making a score of 3.66 on the scale. Hence we derive the following table:

AGE STANDARDS FOR BUCKINGHAM SCALE

Score	Age		Score	Age	
	Years	Months		Years	Months
3.66	9	0	6.46	12	0
3.84	9	6	6.61	12	6
4.36	10	0	7.89	13	0
4.74	10	6	8.15	13	6
5.86	11	0	8.27	14	0
6.13	11	6	8.56	14	6

By interpolating for months, we found the complete table, and from this table we read our scores in terms of problem ages. This same method was used in finding age tables for each scale used. With the spelling scales, we were limited by the fact that the scales do not run continuously through the grades. In the Ayres' list it was necessary to make a separate table for each column, using Ayres' standard per cents for each grade. In Monroe's timed-sentence test, we made a separate table for the test for Grades III and IV, for Grades V and VI, and for Grades VII and VIII.

When we had secured an educational age for each child in each subject tested, we wished to find the arithmetic mean of these ages to procure an average educational age. There were, however, four tests in arithmetic processes, and two in spelling, as against one test in reading and one in problems. We wished our mean to give equal value to skills in each subject. It would be necessary, therefore, to weight the reading age by 4 as a multiplier, the Woody ages by 1, the problem age by 4, and the spelling ages by 2 each in order to equalize their value. This was done in every case except in the two spelling ages. Although we considered spelling ability as important as the other abilities, still we felt that our spelling ages were not reliable enough to warrant equal weighting with the others, and we decided to weight each spelling age by 1 only, thus giving only half value to spelling. Mean educational age, then, is an average of the educational ages in each subject in the following proportion:

	Reading age	=
	Addition age	=
4 ×	Subtraction age	=
	Multiplication age	=
	Division age	=
	Problem age	=
4 ×	List spelling age	=
	Sentence spelling age	=
	Sum of EA's	=
	Divided by 14	= = Mean EA

Educational Ages in C. N. S.

Table I illustrates the record of the chronological, mental, and educational ages which was made for each of the 16 grades. A similar tabulation was put in the hands of the teacher of each grade for reference during promotion week, as summing up for her in an intelligible manner, the results of the educational tests given to her pupils. Compared with the usual method of representing test results, we feel that this summary is far more valuable for diagnosis of individual differences than the usual "list of scores in order" method.

Table I.—A Table Showing the Chronological, Mental, and Educational Ages for Pupils in a Typical VI*B* Grade

Pupils, *girls*	CA	MA	Educational ages								Mean EA
			Reading	Addition	Subtraction	Multiplication	Division	Problem	Sentence spelling	List spelling	
Z. C.	11-3	10-3	11-3	11-4	12-7	14-2	13-5	12	12-0	12-6	12-1
C. C.	12-8	10-4	12-8	11-8	14-0	13-2	12-0	11-0	12-6	12-10	12-2
C. D.	12-10	11-5	11-3	11-10	12-4	13-5	11-10	11-11	12-9	13-0	12-0
D. D.	11-6	11-2	11-3	12-8	14-2	14-2	14-4	12-10	12-9	12-8	11-8
M. E.	10-11	11-9	11-9	11-10	12-10	12-5	14-4	12-6	13-0	13-0	12-6
O. F.	12-3	9-5	13-2	12-0	12-0	13-2	11-8	11-11	10-0	11-0	12-3
S. I.	11-4	10-9	12-8	12-8	13-8	11-10	11-8	11-11	11-6	11-6	12-3
R. K.	10-11	12-5	10-10	14-5	14-0	13-11	15-0	12-6	13-4	13-4	12-8
W. M.	11-4	9-5	11-9	14-9	12-4	13-2	14-4	10-1	11-5	11-10	11-10
G. S.	10-10	11-0	12-8	11-8	12-0	13-11	11-10	11-11	12-2	12-4	12-4
B. W.	12-4	13-6	11-3	13-0	13-8	13-2	11-8	12-6	12-9	12-2	12-3
Median..	11-11	11-1	.....	.....	.....	.....	.....	.....	.....	.....	12-1

With our material now in terms of comparable units, we were able to combine the mental ages and educational ages in one measure—the

Accomplishment Quotient. The Accomplishment Quotient is the ratio of a child's educational age to his mental age; it is the summary of what a child accomplishes educationally compared with what he is capable of accomplishing. In the case of a 10-year-old child who is 12 years mentally, but who does as well in his work as the average 11-year child, his IQ is $^{12}/_{10}$ or 120; his EQ is $^{11}/_{10}$ or 110, but his AQ is $^{11}/_{12}$ or 92. His IQ of 120 shows him to have superior intelligence for his age, but it does not show what he is doing with that intelligence. His EQ of 110 shows that he is doing better work than average for his age, but it does not show that he is doing as well as he is capable of doing. His AQ tells the whole truth—he is accomplishing only 92 per cent as much as he would normally be expected to accomplish, considering his intelligence. We believe that this Accomplishment Quotient is the fairest and most valuable measure now known of the efficiency both of the pupil and the teacher.

The Accomplishment Quotient and the Teacher

From the point of view of the teacher, the Accomplishment Quotient is the only quotient which takes into account the material with which she is working.

1. For the teacher of slow pupils, the AQ interprets fairly the results of her work in relation to the capacities of the children. Her *low* educational scores may appear high when we compare attainment with native capacity.

2. For the teacher of bright pupils, the AQ interprets the educational results fairly. Her *high* scores may not be as high as they should be, when compared with the native ability of the children whom she teaches.

3. For every teacher and every supervisor who is supplied with the AQ's at the beginning of the term, this measure is a protection against injustice. If the class has been improperly taught for 2 or 3 years, the AQ reveals to the teacher and to the supervisors the handicap under which the new teacher is working. If the class has been unusually well taught for 2 or 3 years, the AQ reveals the fact and prevents the new teacher from resting too complacently upon the past laurels of the class.

An illustration from one grade (VIA_2), will demonstrate the unfairness of the "score" method of showing results, and the fairness of our proposed Quotient Method. If we represent the grade medians in

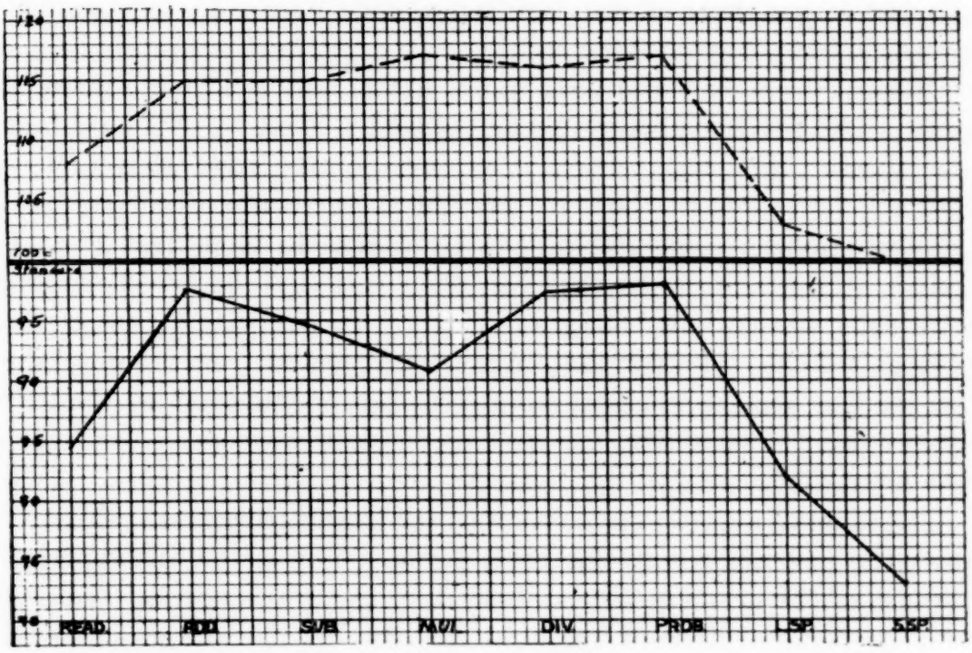


FIG 1.—Results of eight educational tests, Grade VIA_2 when judged by:
Grade medians as per cents of norms (——).
Accomplishment quotients (. . . .).

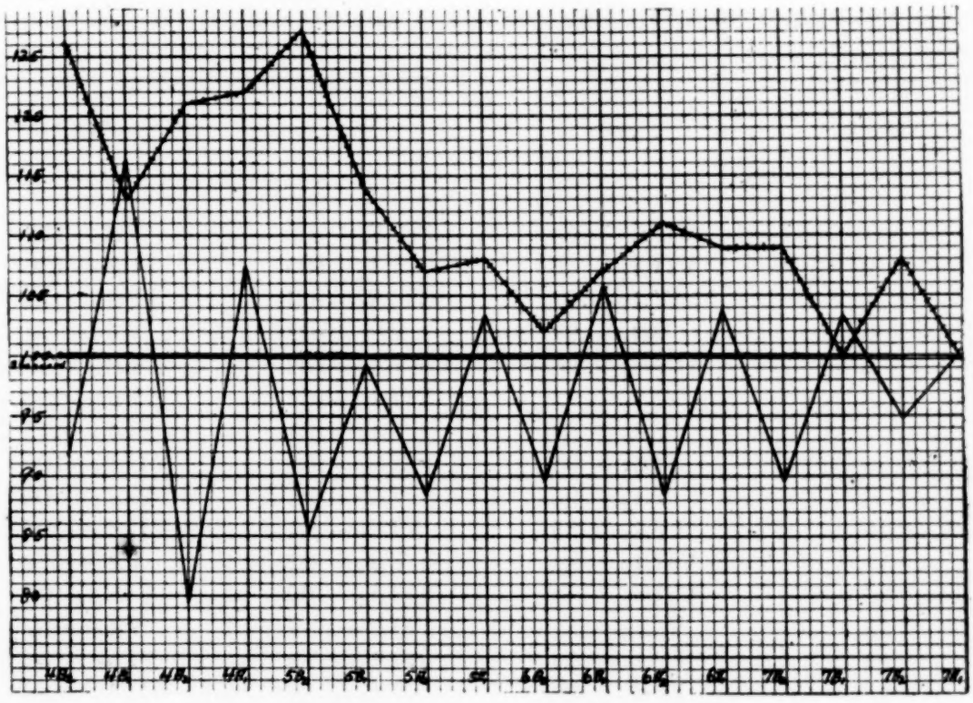


FIG. 2.—Comparison of educational test results in all grades when judged by:
Average grade medians as percepts of standards (——).
Average accomplishment quotients (. . . .).

each test as per cents of standard scores (See Figure 1), the teacher of this grade faces discouraging results. In every test her grade is below the norm. In only three tests is her grade above 95 per cent of the norm. The teacher and the principal realize that the VA_2 grade is a slow grade, but the record does not acknowledge that fact. If however, we represent these results not as per cents of norms, but in terms of median Accomplishment Quotients in each test (*i.e.*, the median of the educational ages in each test divided by the median metal age), notice how this grade shifts. In every test this grade is doing as well as or better than might be expected from its mentality. This teacher should not be discouraged, but encouraged by the excellent progress her children are making in skill subjects in the light of their mentality.

The facts of a grade are consistent for the entire school. Figure 2 shows the average results for each grade. The lower line represents averages of the grade medians divided by the grade norms. In every case, the duller section of the grade presents a discouraging showing. All but one of the brighter sections are above standard in these subjects. But when we take into consideration the mentality of these grades, and represent the results in terms of average

TABLE II.—TABLE SHOWING MEDIAN IQ, EQ, AND AQ FOR ALL GRADES IN RANK ORDER

Rank	IQ	EQ	AQ
1	VIIB_1—103	IVB_1—107	VB_2—127
2	VIIA_1—101	VIA_1—105	IVB_2—126
3	VIA_1— 98	VIIB_1—103	IVA_1—122
4	VB_1— 97	VB_1—102	IVA_2—121
5	IVB_1— 96	IVA_1—102	VB_1—114
6	VIB_1— 94	VIIA_1—101	IVB_1—113
7	VA_1— 94	VA_1—101	VIA_2—111
8	VIB_2— 93	VIB_1—100	VIIB_2—109
9	VA_2— 90	VA_2— 95	VIA_1—109
10	IVA_1— 85	VIB_2— 94	VIIA_2—108
11	VIIA_2— 84	VB_2— 92	VA_1—108
12	VIIB_2— 84	VIIA_2— 90	VIB_1—107
13	VIA_2— 81	VIIB_2— 89	VA_2—107
14	IVA_2— 74	VIA_2— 89	VIB_2—102
15	IVB_2— 72	IVA_2— 89	VIIA_1—100
16	VB_2— 69	IVB_2— 99	VIIB_1—100

Accomplishment Quotients, as shown by the upper line in Figure 2, we find that the teachers who seemingly are doing the best work are those in charge of the slower groups. Two dull sections are achieving over 125 per cent of what might be expected from them. The IVA_2 and the VB_2 grades, the two grades whose medians were lowest in the school, are accomplishing 121 per cent and 127 per cent as much as might be expected from their mental ability and this is the true measure of a grades' achievement—not how much can it score, but how much can it achieve in relation to its ability?

Interesting facts revealed by Figure 2 are that the lower grades seem to be more efficiently taught in these skill subjects than the upper grades, and that all grades in the school are accomplishing as much or more than might normally be expected from their mental abilities.

Table II represents a summary of the median IQ's, EQ's, and AQ's of the 16 grades in order of their rank. The brightest grades tend to have the highest educational age but they tend, after all, to be the laggards, for it is the duller grades that have relatively superior performance. While our Accomplishment Quotients are not an entirely just measure of the teacher's efficiency in these eight phases this year, since they do not take into account with what quotients the grade came to her, they are valuable in revealing the results of 1 year under her teaching, and in another year, we feel a similar record will show very fairly the teaching efficiency of every teacher in these skill subjects.

The Pupil and the Accomplishment Quotient

Educational test results taken alone are no fairer to the child than to the teacher. The bright child receives the high score and the praise; the duller child takes the low score and defeat, with no regard given to the comparative mentalities. The Accomplishment Quotient is a just measure of the pupil's efficiency in school work.

1. For the bright children, it shows which child is living up to his possibilities and which child fails to make his attainment equal to his capacity to attain.

2. For the dull children, it shows which child is needing to be urged and helped still more, which child needs restraining, perhaps, and which are most deserving of praise.

3. Of all children, it asks that the pupil be urged to progress at a rate which is proportional to the mental capacity with which nature endowed him. This is the only fair standard for any child.

Of the 359 pupils tested in the 16 grades, the Accomplishment Quotients are distributed as follows:

Accomplishment Quotient	Number of Pupils
140–150	5
130–139	10
120–129	25
110–119	63
100–109	96
90–99	93
80–89	53
70–79	13
60–69	1
Total.....	359

If we may state that, roughly classifying, an accomplishment quotient of from 90 to 110 is satisfactory, there are, then, about 189 pupils doing satisfactory work, about 103 doing better than might be expected from their native endowment, and about 67 who are apparently not fulfilling their educational possibilities. That is, 52 per cent are doing satisfactory work, 29 per cent are over-attaining, and 19 per cent are falling short of the demands the school may properly lay upon them.

The value of the Accomplishment Quotient in relation to the child is most clearly revealed when we study individual cases. The records of six Grade VII children are reproduced in Table III.

D. M. is an 82 IQ boy, 13 years 8 months of age, doing the work in these subjects of the average child of 12. Instead of censuring the boy for being backward for his age, we see he deserves praise for accomplishing 108 per cent as much as might be expected of him. His profile reveals the fact that his educational work is more than satisfactory in all but reading and spelling, where his educational age lags behind his mental age. Here he needs special attention.

F. Z. is a boy 12 years 8 months of age, with an average educational age of 13 years 6 months. He is doing entirely satisfactory work from the teacher's point of view except for inaccuracy in arithmetic. But his mental age is 16 years 8 months; his accomplishment quotient is only 80. F. Z. is a laggard. Instead of receiving credit for doing well, he needs to be stimulated to do far better work.

E. D. is a normal girl, with an IQ of 101, and an EQ of 99. She is doing satisfactory work for her ability, on the whole.

G. M. is a girl who has been kept after school all her life, and she

TABLE III.—A TABLE TO SHOW THE CHRONOLOGICAL, MENTAL, AND EDUCATIONAL AGES OF SIX VIIA PUPILS SELECTED FOR SHOWING THE INTERPRETATIVE VALUE FOR THE CHILD OF THE SEVERAL MEASURES PRESENTED

Pupil	CA	MA	Educational ages								Quotients		
			Reading	Addition	Subtraction	Multiplication	Division	Problem	Sentence spelling	List spelling	Mean EA	IQ	EQ	AQ
D. M.	13-8	11-2	10-10	12-4	13-3	13-2	13-11	12-10	10-10	10-3	12-0	82	88	108
F. Z.	12-8	16-8	13-11	11-8	12-7	11-5	11-10	14-11	13-9	12-6	13-6	132	107	80
E. D.	13-9	13-11	12-8	15-1	12-10	13-5	13-5	13-9	14-10	14-2	13-6	101	99	97
G. M.	16-9	9-6	11-3	11-0	11-9	11-7	12-0	12-11	10-10	9-9	11-8	57	70	123
S. R.	13-4	15-7	14-4	14-5	14-0	14-6	14-8	13-9	14-8	13-5	14-2	117	106	91
S. F.	10-4	16-0	15-6	15-5	14-4	14-6	14-4	15-6	15-2	16-0	15-3	155	148	95

has faced daily failure and discouragement. Will a teacher ever again scold this faithful child after seeing her accomplishment record? With a mental age of 9 years 6 months, this child is doing the work of an average 11 years 8 months pupil. What does her age of 16 years 9 months matter, when we know that she is accomplishing 123 per cent as much as we might reasonably expect.

S. R. is another case of a bright child, doing supposedly satisfactory work, but really working far beneath her maximum capacity. This girl has never been stirred to do her best; she has been one of the best in the class, and has been praised for doing 91 per cent of what she is capable.

S. F. is a brilliant boy whose unusual intelligence was clearly recognized 2 years ago. In that short time, his educational age has been brought up to 15 years, 3 months, although he is only 10 years 4 months old. He is still accomplishing only 95 per cent of what might be expected from his 16-year intelligence, but it is a fair question whether it is socially wise to give him the educational opportunities sufficient to allow him to make more rapid progress. His present AQ of 95 is a

tribute to what it is possible to do, once the situation is recognized by the teacher.

EA and MA in Classification

The study of individual cases revealed the great need for reclassification in many instances. There was great overlapping of mental ages in the brighter and duller sections of a grade, consequently a conservative reclassification was organized on the following principles:

1. A child whose MA is above the median MA of the next grade should be considered for double promotion.
2. Of these, only those children whose EA's are above the median EA of their present grade should be allowed a double promotion, since without the necessary tools for acquiring the work of the omitted half-year, the child could hardly hope to succeed.
3. A child whose MA is below the median MA of the grade below should be considered for non-promotion.
4. Of these, only those children whose EA's are below the median EA of their present grade should be non-promoted.
5. All other children should be regularly promoted, and assigned to the brighter or duller group on the basis of test results.

It is the opinion of the writers, however, that group test results must always be subjected to the judgment of the teacher for corroboration, and where disagreement is sufficiently marked to cause a possibility of injustice to the child, individual case study must follow. In no other way can we avoid letting cold figures work disastrously in the career of a living child.

In conclusion, may we give a statement of our faith? We believe that the Accomplishment Quotient is the fairest and most valuable measure of both the efficiency of the teacher and the pupil, that by reliance on it for guidance, the teacher will come to exact from him that hath even more than he has been giving and take from him that hath not even less than he has been able to give. And the educational plaudits of "well done" are seen to be merited more by the retardate possessing the one than by the accelerate endowed with the ten.

THE TEACHING OF EDUCATIONAL PSYCHOLOGY IN THE UNITED STATES

H. H. REMMERS

Colorado College

AND

F. B. KNIGHT

University of Iowa

No one any longer thinks, seriously or otherwise, of describing psychology, and particularly educational psychology, as "putting what everybody knows in language which nobody can understand."[1] The viewpoint of Pyle as expressed in his recent book is probably nearer the concensus of opinion among those qualified to judge in the matter. "Educational psychology," he says, "is an experimental science,"[2] and it will be our purpose so to treat it in what follows.

To determine the extent to which experiment is used in elementary courses of educational psychology has been the object of the study to be described in this chapter. What is the present practice? To what extent do the beginning courses in educational psychology—which represent the sum total of the training in psychology that all but a negligible part of young teachers in training receive—attempt to give the prospective young teachers a mastery of the subject through an experimental approach? Such a mastery, that is, that will give her the *savoir faire* which comes only from the ability to apply theory directly and accurately to classroom problems?

It might be contended, if it did not assume the reader's consent without argument, that the human mind induces a theory from much practice rather than that it deduces current practice from theory. An order of merit of learning techniques might be roughly:

Best—doing the thing oneself.
Next—seeing it done.
Third—reading about what happened when it was done.
Last—reading about the theory that underlies the operation.

Thus many a good teacher takes pains to make the assignments clear, exact, and definite without knowing that "ease of identification of

---

[1] Welton, J.: "The Psychology of Education," (1912), p. 1.
[2] Pyle, W. H.: "The Psychology of Learning," (1921), preface.

bonds" is helpful in learning; conversely, many a conscientious student of educational psychology knows in a verbal fashion that "ease of identification of bonds" is an aid to learning, but never sees that it means something in the practical situation of assigning lessons.

Further, what are considered, for the elementary student, to be the most profitable experiments? Answers to these questions were obtained through correspondence with the leading universities, colleges, normal, and teachers' training schools throughout the United States. The letter sent is here inserted:

DEAR SIR:

In a study that is being conducted in the Graduate College of the State University of Iowa, we are attempting to gain a true picture of the place of experiment in *elementary* or *first courses* of educational psychology.

The following meanings are attached to the terms as used: (1) An *experiment* is that which is done by the students themselves singly or in groups; (2) a *demonstration* gives the students an opportunity to observe an experiment as carried out by the instructor and assistants; (3) a *recountal* of experimental data places emphasis in lectures on experiments carried out by research workers.

We shall greatly appreciate your cooperation in this study, and will reciprocate by sending you a report of our findings if you wish it. Please place a check-mark (✓) in the appropriate column after the items in the accompanying list thus indicating what you are offering in your courses. Also please answer the three questions at the end of the list.

Very truly yours,

—— ——.

Note: These experiments have purposely been left disorganized, since no one organization would correspond to the sequence of topics in all courses.

	Experiment	Demonstration	Recountal
1. Analyzing bits of behavior for presence of original nature			
2. Building up learning curves			
3. Mirror drawing			
4. Building up forgetting curves			
5. Testing for individual differences (opposites test, etc.)			
6. Silent reading tests			
7. Arithmetic tests			
8. Experiments on sensation			
9. Waxing and waning of original tendencies			
10. Tests for visual defects			
11. Tests for auditory defects			
12. Trial and error learning			
13. Transfer of training			
14. The work curve to show entrance of fatigue			
15. Illustration of laws of attention:			
(*a*) Effect of intensity			
(*b*) Effect of contrast			
(*c*) Counter-attraction, etc.			
16. Memory; part vs. whole methods			
17. Statistical method:			
(*a*) Central tendencies			
(*b*) Distributions			
(*c*) Correlations			
18. Introspective questionnaire on vividness of imagination			
19. Test for immediate memory such as Knox cube			
20. Apperception:			
(*a*) Interpreting ink blots			
(*b*) Hidden pictures			
(*c*) Word completion			
21. Fluctuation of attention			
22. Ergograph			
23. Experiments on testimony and report			
24. Suggestibility			
25. Normal associations vs. abnormal associations			
26. Eye-movements in reading			
27. Bodily changes in emotion			
28. Pleasantness vs. unpleasantness in color combinations			
29. Testing for IQ			
30. Testing for special abilities			
31. Testing for educational accomplishment			
32. Testing for emotional stability			
33. Associative shifting			
34. Influence of drill			
35. Finding plateaus in learning curves			
36. Questionnaire on children's interests amusements, ambitions, ideals			
37. Puzzles in illustrating trial and error chance variations, multiple response, etc			
38. Motion picture			
39. Survey of play activity			
40. Psychological analysis of misbehavior types			
41. Experiments in animal psychology to illustrate habit formation, situation-connection-response-series			

(*A*) Please list any others not listed above that you have found valuable.

(*B*) Assuming you have reasonable conditions and time for giving your first course in educational psychology please star the experiments which you consider of unquestioned value for relatively immature students.

(*C*) Please list texts and manuals used in your beginning courses.

Name of your institution.....

The method of constructing the foregoing letter was that of a rather thorough canvass of all the textbooks and experimental manuals on the market. From these were selected the experiments listed, as

being fairly representative of the general field of experimentation on the subject.

Forty-one returns were obtained—a fair sample, we believe, of the more progressive institutions throughout the country.[1]

---

[1] The names of these institutions follow:
Cornell University, Ithaca, N. Y.
Dartmouth College, Hanover, N. H.
Drake University, Des Moines, Iowa.
Eastern Illinois State Teachers College, Charleston, Ill.
George Peabody College for Teachers, Nashville, Tenn.
Harvard University, Cambridge, Mass.
Illinois State Normal University, Normal, Ill.
Iowa State College, Ames, Iowa.
Iowa State Teachers College, Cedar Falls, Iowa.
Montclair State Normal School, Montclair, N. J.
New York State College for Teachers, Albany, N. Y.
Northeast State Teachers College, Kirksville, Mo.
Oberlin College, Oberlin, Ohio.
Ohio Wesleyan University, Delaware, Ohio.
Pennsylvania State College, State College, Pa.
Purdue University, Lafayette, Ind.
Rutgers College, New Brunswick, N. J.
Sophia Newcomb Memorial College, New Orleans, La.
Southern Branch University of California, Los Angeles, Cal.
State Normal College, Ypsilanti, Mich.
State Normal School, Indiana, Pa.
State Normal School, Lewiston, Idaho.
State Normal School, Fitchburg, Mass.
State Normal School, Stevens Point, Wis.
State University of Montana, Missoula, Mont.
Teachers' College, Kearney, Neb.
Teachers College, New York, N. Y.
Tufts College, Mass.
University of Arkansas, Fayetteville, Ark.
University of California, Berkeley, Cal.
University of Georgia, Athens, Ga.
University of Kansas, Lawrence, Kansas.
University of Maine, Orono, Me.
University of Oklahoma, Norman, Okla.
University of Rochester, Rochester, N. Y.
University of Texas, Austin, Texas.
University of Virginia, University, Va.
University of Washington, Seattle, Wash.
University of Wisconsin, Madison, Wis.
Western State Normal School, Kalamazoo, Mich.
Yale University, New Haven, Conn.

Turning now to the interpretation of the data, let us note a tabulation of the items checked in the respective columns exclusive of "Recountal."

TABLE I[1]

	Experiments	Demonstration	Experiment starred		Experiments	Demonstration	Experiment[2] starred
1	10	12	1	22	2	3	0
2	28	17	18	23	14	8	10
3	18	7	7	24	8	7	0
4	13	10	6	25	6	8	1
5	29	13	15	26	7	9	1
6	20	11	8	27	2	6	1
7	15	12	6	28	9	7	0
8	18	4	5	29	19	19	13
9	0	2	0	30	13	9	8
10	14	18	6	31	15	9	7
11	10	16	5	32	4	4	1
12	20	12	9	33	3	1	1
13	10	7	8	34	10	6	8
14	6	6	1	35	19	7	8
15	17	8	9	36	4	3	1
16	25	10	14	37	8	8	3
17	28	14	11	38	1	2	1
18	21	8	5	39	2	4	3
19	15	10	1	40	1	2	0
20	23	9	7	41	2	5	3
21	14	11	3				

It will be noticed that the following experiments stand out as the most frequently checked both under "Experiments" and "Experiments starred:"

2. Building up learning curves
5. Testing for individual differences
16. Memory—part vs. whole methods
17. Statistical method
29. Testing for IQ. Less, though relatively high importance is given under "Experiments starred" to

[1] The consecutive numbers 1, 2, 3, . . ., 41 correspond to items in the circular letter.

[2] "Experiments starred" refers to answers to question *B* in the circular letter.

13. Transfer of training
23. Experiments on testimony and report
34. Influence of drill
35. Finding plateaus in learning curves.

It may be argued from this that the laboratory part of a beginning course in educational psychology should have these experiments for a nucleus, to which additions might be made as circumstances and available time may allow.

There is, however, a disappointing lack of unanimity of opinion apparent in the foregoing tabulation, both as to what is offered in the laboratory and what is considered of importance for elementary students.

In answer to Question *C* of the letter—a request to list texts and manuals used—the respondents did not in every case give the name of the book, but merely the author. We have therefore thought it best to give only the author's name as representing a general viewpoint of his relative influence, and the corresponding frequency of mention.

Author	Frequency
Angell	2
Averill	3
Bagley	2
Baldwin	1
Betts	2
Breese	1
Bolton	1
Breitweiser	2
Calkins	1
Colvin	7
Dewey	1
Freeman	5
Gordon	2
Halleck	1
Hunter	1
Hollingworth and Poffenberger	1
James	1
Judd	1
Kirkpatrick	2
Langfeld and Allport	1
LaRue	2
Monroe	4
Norseworthy and Witley	2
Ogden	1

Author	Frequency
Parker	1
Peterson, H. A	1
Pillsbury	5
Seashore, C. E	6
Starch	11
Strayer	2
Strong	2
Terman	3
Tracy	1
Thorndike	10
Waddle	1
Watson, John B	1
Woodworth	4

Unless it be argued that an introductory course in general psychology furnishes an adequate psychological background for the teacher, it is obvious that the texts of many of the authors listed hardly fit into a course in educational psychology. They are merely texts adapted for general beginning courses, and the feeling that they are inadequate for the specific purposes of elementary courses in educational psychology is corroborated by the relatively large number of textbooks in educational psychology published within recent years. Ideally, the best way to train the teacher is "on the job." Courses in practice teaching recognize this fact. It is only from consideration of economy and feasibility that all teachers are not so trained. The principle of as direct application of theory to practical problems as possible holds in psychology as well as in teaching problems in general, and the notion that general psychology is sufficient is only one step removed from the much-belabored idea of formal discipline.

The teacher must have not merely a theoretical but primarily a Socratic knowledge of psychology; it must function in the common, every-day come and go of the classroom. There is reason to suspect that too often the prospective teacher's information on matters psychological operates only for purposes of regurgitating for the professor a more or less organized body of abstractions, which are to be relegated—with a sigh of relief—to the mental ash heap immediately after the final examinations. Much the same argument holds concerning several of the experiments listed in the questionnaire and those given in answer to Question *A*.

The following are some of the obviously inept experiments so listed:

Weber's law
Reaction time
Ampullar sense (revolving table)
Threshold differences—skin
Laws of mixture, contrast, and adaptation
Conditions of spatial perception
Study of dreams
Clairvoyance, mediums, automatisms

The thesis here contended for is not that this does not represent useful knowledge, but simply that it is not the most useful knowledge for young teachers in training, who in any case can give but little time to the study of psychology.

The lack of unanimity of opinion concerning what should go into a beginning course in educational psychology is apparent not only in the foregoing, but in the texts that deal directly with educational psychology as well. An attempted analysis of the contents of five of these reveals a decided divergence in point of view, organization, and content; so much so, in fact, that a projected comparison of them here had to be abandoned. The reader may check on this statement by comparing the following:

Starch, D.: Educational Psychology.
Averill, L. A.: Psychology for Normal Schools.
Gordon, Kate: Educational Psychology.
Cameron, E. H.: Psychology and the School.
La Rue, D. W.: Psychology for Teachers.

Making due allowance for the individuality of the author, it must still be true that there is a best way of presenting the subject, as well as a best way of teaching it. What this best way is, yet remains a problem to be solved.

Differences in Viewpoint

That there are some fundamental differences in viewpoint may further be seen by comparing that of Starch and Thorndike, who as indicated by the returns are the leaders in educational psychology. Starch[1] definitely belittles the place of instincts in educational practice. "The direct appeal to, and use of, instinctive reactions in actual concrete instances in school work," he asserts, "are not as frequent and specific as is commonly implied." That this opinion is hardly shared by Thorndike is evidenced by the fact that he devotes the whole of Volume I of his three volume work in Educational Psychology to a discussion of original tendencies.

[1] Starch, D.: "Educational Psychology." 1920, p. 12.

The issue is clearly drawn in a review[1] of Starch's book. "He" (Starch), says the reviewer, "points out that instincts as such have very little significance for education, and that the chief educational doctrines based upon instincts (dynamic theory of instincts, transitoriness, and recapitulation) have very little justification in verified fact. This is a sane view, and we hope that it will tend to neutralize the overemphasis of instincts that has been prevalent in educational discussion since James." It is probably safe to assume that the advertising expert makes his appeal to the same sort of individual, psychologically, as does the educator. Yet the theory of the psychology of advertising is in great part told when instinctive trends have been thoroughly discussed.

Such fundamentally different attitudes as held by Starch and Thorndike cannot both be right. And educational practice cannot be the same whether we hold to the one or the other. If original nature has relatively little weight in the scales, then it follows that a correspondingly larger amount of nurture (*i.e.*, education) will be required to produce the socially efficient individual. If we think of man as a product of original nature and education, the equation:

$$X \times Y = Z$$

would seem to express this relation when X equals original nature, Y equal nurture, and Z equals the end product—man. It is obvious that to have Z remain a constant when either X or Y are changed in value requires a corresponding change in the other member.

Conclusions

1. There is considerable evidence to show that educational psychology is definitely regarded as an experimental science.

2. There is, with a few exceptions, little unanimity of opinion as to what is most important for elementary laboratory courses in the science.

3. Concerning the importance of a few items, as individual differences, statistical method, learning curves, experiments on memory, and testing for intelligence, there is relatively high agreement.

4. A survey of textbooks shows that there is little agreement as to point of view, organization, and content.

5. The two authors most influential in the United States in educational psychology—Starch and Thorndike—disagree fundamentally as to the importance of original nature in the science.

---

[1] *Jour. Educ. Psy.*, Dec., 1920, pp. 535–6.

A COMPARISON OF MENTAL AGE SCORES OBTAINED BY PERFORMANCE TESTS AND THE STANFORD REVISION OF THE BINET-SIMON SCALE

BUFORD JOHNSON

AND

LOUISE SCHRIEFER

Bureau of Educational Experiments, New York City

The use of non-verbal tests in the determination of age differences in mental traits is recognized as especially important in the early years. The selection of a series of such tests that will differentiate throughout a period of years is still an experimental problem. In the systematic study from year to year of the various age groups in the city and country school we have included the Pintner-Paterson Performance Scale and the Stanford-Revision of the Binet-Simon Scale. This study reports a statistical evaluation of the scores made by the same children when measured by both scales. The children studied were so young both scales could be given during one laboratory examination. No discussion is given of the changes in scores for the individual from year to year. That forms another study which is in progress. All the scores considered were obtained during the year 1920–21. Some of the children had been in the school for several years and had been tested each year. Forty-one children entered the school during this year and had only the one experience with either group of tests. The correlation of each type of scale with chronological age is determined. Scores in other non-verbal tests are also presented and included in the comparative study. These tests do not form the complete series given to these children but from the records we have taken the scores made in the following tests for the various age groups as a basis for this study.

Three-year Group.—Stanford Revision of the Binet; Manikin, Seguin, Mare and Foal of the Pintner-Paterson Performance Scale; the Witmer Cylinder Test; Rossolimo's Pictures No. 1–4.

Four- and Five-year Groups.—Stanford Revision of the Binet; Mare and Foal, Seguin, 5-figure, 2-figure, Casuist, of the Pintner-Paterson Performance Scale; the Witmer Cylinder Test; Action Agent; Rossolimo's Pictures No. 1–7.

Six-year Group.—Stanford Revision of the Binet; Mare and Foal,

Seguin, 5-figure, 2-figure, Casuist, Diagonal, Triangle, Ship, Feature Profile, Knox Cubes of the Pintner-Paterson Performance Scale; the Witmer Cylinder Test; Action Agent: Rossolimo's Pictures No. 1–8.

Seven-year Group.—The six-year tests and in addition Healy *A*; Substitution of the Pintner-Paterson Performance Scale; Rossolimo's Pictures No. 1–10; Dearborn Formboard.

Eight- and Nine-year Groups.—Stanford Revision of the Binet: Complete Pintner-Paterson Performance Scale with the exception of the Adaptation Board; Healy Picture Completion II.

The Method.—The directions for giving and for scoring these tests may be found in the following publications: Terman: "The Measurement of Intelligence," Houghton Mifflin; Pintner-Paterson: "A Scale of Performance Tests," Appleton; Bureau of Educational Experiments: "Health Education and the Nutrition Class," E. P. Dutton.

The group of 86 children is made up of both boys and girls ranging in age from 3 to 9.8 years. The age groups are formed on the basis of the age in years that each child had reached at the time of the testing. Children in the 3-year group had reached their third birthday but had not yet reached their fourth; consequently the theoretical average for each age group falls at 3.5, 4.5 years, etc.

Although the Jewish and Italian nationalities are represented, it is primarily an American born group, its members coming from families widely variant as to fields of activities, professional and industrial. Including mothers and fathers, we find 90 parents are represented. The number of children in a family range from one to four. The classification according to vocation shows that 36 per cent are in the professions; 17 per cent are artists; 37 per cent are in commercial occupations, and 10 per cent are in miscellaneous activities, such as teamster and detective. It seems a representative group of city children, many of whom were born outside of New York City. Admission to the school is not wholly dependent upon ability to pay the tuition fees as there are some scholarships available. Other factors that enter into the selection of the group, aside from the exclusion of children below the normal rating in intelligence, are the possibilities of continued attendance of the same children throughout a period of years, and the cooperation of the parents in an educational scheme involving a research program where such cooperation is especially desirable.

The Stanford Revision IQ's range from 96–167, with an average of

114.3 s.d. 11.3. This average places the group in the superior intelligence class as rated by Terman. The distribution is as follows.

IQ	Per Cent
96–105	17 or 20
106–115	37 or 43
116–125	18 or 21
126–135	12 or 14
136–145	1 or 1
146–155	—
156–165	—
166–175	1 or 1
	86 100

In the Terman study of 905 children between 5–14 years of age, 69 per cent of the children obtained an IQ of 96 or more; in the City and Country School 100 per cent of the children obtained an IQ of 96 or more, 25 per cent being in the very superior intelligence class according to the Terman rating, that is, IQ 120 or more. We do not believe this is a very unusual distribution for children of these ages in private schools. The young child in a stimulating environment has an intelligence quotient as determined by the Stanford Revision that we do not expect to be constant but to decrease with increasing years. Slight variations in scores make greater differences in intelligence quotients than are possible in later years.

The Pintner-Paterson Tests were not standardized for the lower ages so that it was necessary for us to determine median age limits from our own test results in order that a rating might be made of the younger children. According to the method described in the Pintner-Paterson Scale (p. 151) for determining the limiting values for various ages we have determined 3-year and 4-year limits for Seguin and Mare and Foal, using the available Pintner-Paterson norms and modifying our own where over-lapping occurred; that is, Pintner-Paterson have set their lower 5-year limit in time for Seguin Formboard at 50 seconds. Our lower limit for 4-years would fall at 44 seconds, obviously conflicting with the Pintner-Paterson 5-year norm. We have, therefore, modified our 4-year limits and used the following:

	3-year	4-year	5-year
Seguin time in seconds.....	110– 55	54– 51	50–32
	300–226	225–151	150–89

These test values have been determined for only a few cases including eight 3-year-old children and fifteen 4-year-olds.

The following table contains the averages of chronological age, of Stanford Revision Mental Age, and the Pintner-Paterson Median Age for each of the year groups:

Age group	Number of cases	Chronological age in months		Terman mental age in months		Pintner-Paterson median age in months	
		Average	σ	Average	σ	Average	σ
3	8	40.5	4.27	48.0	6.63	58.5	13.99
4	15	53.5	3.75	60.4	5.76	64.8	17.83
5	21	66.3	3.32	76.9	8.52	89.9	25.42
6	21	75.7	2.90	86.7	6.07	109.6	19.64
7	7	89.3	2.62	104.4	10.78	143.1	29.13
8	5	102.4	3.38	108.4	4.96	134.4	19.20
9	9	115.0	2.83	129.8	12.57	146.0	17.89

Computing the differences between the Stanford Revision and Pintner-Paterson measures for each year group and relating them to the probable error of their differences, we find a significant difference for each age except the 4-year, which shows only a possible difference. The largest difference occurs in the 6-year group, the next largest at 7 and 8, which again suggests the greater usefulness of the Performance scale at the ages of 6, 7, and 8 years.

PROBABLE ERRORS OF DIFFERENCE BETWEEN PINTNER-PATERSON AND TERMAN, RELATED TO THE ACTUAL DIFFERENCE

AGE GROUP	$\frac{D}{PED}$
3	$\frac{10.5}{3.68} = 2.85$
4	$\frac{4.4}{3.25} = 1.35$
5	$\frac{13.00}{3.94} = 3.29$
6	$\frac{22.9}{3.02} = 7.58$
7	$\frac{38.7}{7.92} = 4.88$
8	$\frac{26.0}{5.99} = 4.34$
9	$\frac{16.2}{4.91} = 3.30$

There are 22 children having a Stanford Revision IQ of 120 or more. These would be considered in the very superior group of intelligence. They distribute themselves as follows:

Age.....	III	IV	V	VI	VII	VIII	IX
	3	2	6	6	2	0	3

and comprise one-fourth of the entire group. The distribution shows that approximately one-third of each of the separate age groups is included, except in the Four Group, where only 2 out of 15 children have very superior intelligence according to the Stanford Revision rating, and in the Eight Group, where there are none.

Comparing these Stanford Revision IQ with IQ's similarly calculated for the Pintner-Paterson Median Age, we find that 19 of the 22 have Pintner-Paterson IQ's of 150 or more. The remaining 3 IQ's are 140, 133, 133, which in ordinary selection would be considered a high rating. This similarity in scores suggests that children of high general intelligence as measured by the Stanford Revision scale are also high in performance tests.

Upon the suggestion of Pintner and Paterson (p. 157) that the children above the middle 50 per cent of a given troup are probably bright, we have selected from each of our age groups the children of the upper 25 per cent in Pintner-Paterson Rating; this forms a basis for comparison with the very superior group according to the Stanford Revision Rating. The following table shows the distribution of this grouping; obviously. it includes approximately one-fourth of the entire group.

PINTNER-PATERSON MEDIAN AGE DISTRIBUTION FOR UPPER 25 PER CENT OF EACH YEAR GROUP

Year group	Median age								
	6	7	8	9	10	11	12	13	14
3	1	1							
4	..	2	2						
5	..	..	..	2	1	..	2		
6	..	..	..	..	..	3	2		
7	..	..	..	..	..	..	..	..	2
8	..	..	..	..	..	..	..	..	1
9	..	..	..	..	..	..	..	..	2

Comparing these ratings with the Stanford Revision ratings of the same children, we do not find the agreement that was noted above. Only eight of these children would be included in the very superior

intelligence class according to the Stanford Revision. The remaining 13 are distributed as follows:

Stanford Revision IQ....	95–99	100–104	105–109	110–114	115–119
	2	..	5	4	2

The two children below 100 in Stanford Revision IQ are in the 7- and 9-year groups. Analyzing their performance in individual tests, we find the 7-year old considerably above his class median in 6 out of 8 tests; in 2-figure time the class median was 45, the individual's time 65 seconds; in Knox Cubes the median score was 6, the individual's score 5. In the Stanford Revision he succeeded in the reproduction of familiar things, such as counting backwards, the date, easy definitions, and in weight discrimination. He failed, however, in tests requiring language expression and abstractions, *i.e.*, comprehension, similarities, sentences, rhymes, vocabulary. The 9-year boy falls below his group median in only one test of 14; in 8 tests he is above the median. Even in the Knox Cube test, which is primarily a memory test, and in Picture Completion and Substitution, which involve

TABLE I.—MEDIAN AGE FOR INDIVIDUAL TEST SCORES

Year group	Mare and Foal time	Seguin time	5-figure time	5-figure errors	2-figure time	2-figure moves	Casuist time	Casuist errors	Triangle time	Triangle errors	Diagonal time	Diagonal errors	Healy A time	Healy A moves	Manikin score	Feature time	Ship score	Picture completion	Substitution	Cube test	Group median age
3	5	3	..	....	..	....	..	..	..	..	..	..	..	..	6	..	....	....	....	..	5.0
4	6	5	6	6.0	6	7.0	6	..	..	..	..	..	..	..	7	..	....	....	....	..	6.0
5	7	5	7	10.0	7	8.0	7	10	10	15	6	13	..	..	7	..	5.0	....	....	7	7.0
6	..	6	8	12.5	8	9.0	8	12	7	15	7	13	..	..	..	15	7.0	....	....	7	8.0
7	..	..	13	15.0	10	12.5	12	15	9	15	12	14	7	12	..	..	12.0	....	....	9	12.0
8	10	10	12	15.0	10	10.0	13	12	11	15	14	14	7	15	8	12	9.5	9	10	12	11.5
9	13	12	14	15.0	11	10.0	10	12	13	15	11	12	7	15	8	15	12.0	11	10	14	12

complex mental processes, the child runs ahead of his class median. In the Stanford Revision his performance was similar to that of the 7-year old, in that he passed tests involving reproduction and memory such as word naming in the 10-year old series, and the repetition of digits backward in the 12-year series; he gave a fairly creditable performance in Comprehension of Questions. The tendency is to pass tests involving construction from concrete materials but to fail in tests emphasizing comprehension and abstract reasoning as the Stanford Revision tests do.

These results again suggest that superior general intelligence is accompanied by a high degree of ability in performance. In medium levels this is not consistently true. A high performance rating may be made by one who shows mediocre ability in general intelligence tests.

In Table I we have tabulated the median ages which correspond to the median test scores obtained by the children of our group. These individual medians have been summarized in the last column under "Group Median Age." It is evident that each of the year groups, 3, 4, 5, 6, are 2 years ahead of the Pintner-Paterson standards, the 3-year group having a 5-year rating, the 4-year group a 6-year rating, etc. The 7-year group has a median age of 12, which is undoubtedly too high to be accurate, but the interesting fact is that the scale fails to differentiate between the 7-, 8-, and 9-year group. Some differentiation of course would be expected since 7 tests have 14-year old norms and 3 tests have 15-year-old norms.

Continuing the method referred to above of determining median age limits, in Table II we present these limits for our group. They are only approximations, owing to an arbitrary determination of the limits where intermediate ages were missing. The lower limits in such cases have been determined by the lower quartile of a given age, the upper limits by the upper quartile. They are open to criticism and are presented merely for their suggestive value. Substitution, Healy A, Picture Completion and Knox Cubes have been omitted

TABLE II.—MEDIAN INTERVALS FOR SCORING TESTS
Determined by results obtained in the city and country school

Year group	3	4	5	6	7	8	9
Mare and Foal	170–91	90–66	65–50	.....	.....	49–33	32–0
Seguin	110–55	54–44	43–33	32–31	.....	17–15	14–0
Manikin	3	4	4	.....	.....	5	5
5-figure time	.....	300–181	180–125	124–79	78–53	52–43	42–0
2-figure time	.....	300–146	145–97	96–61	60–45	44–43	42–0
Casuist time	.....	30–238	237–146	145–85	84–64	.....	63–0
Triangle time	.....	.....	.....	120–67	66–54	53–40	39–0
Diagonal time	.....	.....	300–98	97–56	55–25	24– 0	
Ship score	.....	.....	6–12	13–18	19–20		

because only a very limited number of measures were available; errors and moves in 5-figure, 2-figure, Casuist, Triangle, and Diagonal showed very little differentiation.

TABLE III

		Chronological age		Terman mental age		Pintner-Paterson median age	
		Number of cases	Correlation coefficient	Number of cases	Correlation coefficient	Number of cases	Correlation coefficient
Mare and Foal.....	Time	53	0.554	53	0.584	53	0.582
Seguin.....	Time	63	0.650	63	0.688	63	0.504
Five figure board...	Time	65	0.586	65	0.603	65	0.697
Five figure board...	Errors	..	.....	61	0.401		
Two figure board...	Time	60	0.579	60	0.335	60	0.441
Two figure board...	Errors	..	.....	58	0.200		
Casuist.....	Time	64	0.503	64	0.536	64	0.516
Casuist.....	Errors	..	.....	63	0.203		
Triangle.....	Time	60	0.298	60	0.335	60	0.387
Triangle.....	Errors	..	.....	58	0.120		
Diagonal.....	Time	51	0.212	51	0.289	51	0.592
Diagonal.....	Errors	..	.....	50	0.217		
Healy Puzzle *A*.....	Time	16	0.040	16	−0.205	16	0.131
Healy Puzzle *A*.....	Moves	..	.....	16	0.249		
Manikin.....	Score	52	0.533	52	0.569	52	0.657
Feature profile.....	Time	25	0.270	25	0.267	25	0.385
Ship.....	Score	57	0.534	57	0.952	57	0.632
Picture completion..	Score	13	0.548	13	0.243	13	0.532
Substitution.....	Score	13	0.174	13	0.230	13	0.488
Knox cubes.....	Score	56	0.517	56	0.536	56	0.546
Cylinders.....	Time	86	0.565	86	0.576	86	0.669
Rossolimo.....	Time	80	0.264	80	0.504	80	0.340
Rossolimo.....	Score	83	0.812	83	0.836	83	0.853
Action agent.....	Time	62	0.360	62	0.367	62	0.292
Action agent.....	Score	62	0.455	62	0.473	62	0.349
Healy picture completion II.....	Score	13	−0.016	13	0.204	13	−0.147
Dearborn II.....	Time	19	0.218	19	0.400	19	0.354
	Moves	..	.....	19	0.083		
Chronological age...	.....	..	.....	86	0.946	86	0.782
Terman mental age.	.....	..	.....	..	.....	86	0.818

The curves for time measures suggest that time is a good measure up to a certain point in a child's development; having attained a certain limit, however, it is rarely possible for an individual as he increases his age to lessen the time of a given performance. If he can lessen it, practice and exceptional motivation have probably made it a special ability.

The correlations with age show more plainly this relationship between age and a given test. The following correlations are arranged in rank order:

Chronological age with	Seguin	0.650
	5-figure time	0.586
	2-figure time	0.579
	Mare and Foal time	0.554
	Picture completion score	0.548
	Ship score	0.534 (limited number)
	Manikin score	0.533
	Knox score	0.517
	Casuist time	0.503

There is a sharp line of demarcation between the above mentioned tests and the remaining ones that arranged in rank order are:

Chronological age with	Triangle time	0.298
	Profile time	0.270
	Diagonal time	0.212
	Substitution score	0.174
	Healy a time	0.040

Discussion and Summary

Judging from fineness of discrimination, age progression and correlation, the following tests prove themselves most adequate; Mare and Foal, Seguin, 5-figure, 2-figure, Casuist, Manikin, Ship, and Cubes. Of the remaining tests, Triangle and Diagonal are fairly good. Picture Completion and Substitution are untried because of the limited number of measures, but the indications are that they would fall in the desirable group. These tests correspond closely to the Pintner-Paterson Short Scale which includes Mare and Foal, Seguin, 5-figure, 2-figure, Casuist, Manikin, Feature and Ship, Picture Completion and Cubes. We would omit, however, from the Performance Scale "errors" and "moves" as measures on the basis of lack of age progression and differentiation between the ages, also difficulty of recording.

In addition to the Stanford Revision and Pintner-Paterson Ratings,

we obtained a score in the Dearborn Group tests of Intelligence, Series I. The average score in Dearborn was 73, this being the norm for 10½-year old children. Our children average 9.2 years at time of testing.

Dearborn correlated 0.55 with Stanford Revision Mental Age and 0.83 with Pintner-Paterson Median Age. The high correlation between Dearborn and Pintner-Paterson would be expected from the fact that language expression plays no part in either of them, the chief requirement being to execute directions quickly and accurately. It would be interesting to follow up the relationship between the two measures, because of the time saving involved in giving the Dearborn Test instead of the Pintner-Paterson Scale. The correlation between Dearborn and Stanford Revision is not high.

The Stanford Revision Scale correlates more highly with Chronological Age than with the Performance Scale, but correlates more highly with the Performance Scale than does the Performance Scale with Chronological Age.

Specific tests in the Performance Scale have fairly high correlation-coefficients with Chronological Age and with the Mental Age scores obtained by both scales.

Those who made high intelligence quotients by the Stanford Revision make high ratings in Performance Tests. There are some who make high Performance Scale ratings but hardly attain average Intelligence Quotients.

The Rossolimo series of graded pictures is the best single performance test as measured by high correlation with the Stanford Revision Scale.

It is indicated that a series of Performance Tests may be selected which will give ratings comparable with the Intelligence Quotients and will show good age progression. This series is especially desirable for use with young children in order that we may supplement data which are obtained by methods that put a premium upon vocabulary attainments.

CORRELATIONS OF FOUR INTELLIGENCE TESTS WITH GRADES

A. M. JORDAN

University of Arkansas

There are two ways of progressing in the development of mental tests: The first is to invent new tests, while the second is to refine and improve those extant. In order to develop or refine a test, knowledge of its advantages and limitations is necessary. This knowledge is gained by correlating the test as a whole, and the several elements of the test with various criteria. From the criteria which might be used the one chosen in this study consists of the grades made by the students in their subjects of instruction. The reliability coefficient of grades in successive semesters is around 0.75 so that the criterion chosen varies.

The purpose of the investigation is: First, to find out the group test of the elements of the group test which furnish the best prognosis of the standing of pupils in the high school subjects of instruction when the total grades are combined into one score. The second purpose is to discover the test which correlates most highly with the grades received in English, mathematics, general science and history.

The method employed was: First, to give four intelligence tests (Army Alpha, Terman, Otis, and Miller) to 67 high school pupils;[1] second, to give the Army Alpha to 315 university students. With this material collected the next task was to proceed with the correlations. Correlations together with probable errors were computed between (*a*) each of the four group tests given to high school pupils and the total points of the subjects of instruction. The value of each subject was determined by grade points: $A = 6$, $A-$ and $B+ = 5$, $B = 4$, $B-$ and $C+ = 3$, $C = 2$, $C-$ and $D+ = 1$, $D = 0$, $E = -1$, and $F = -2$. (*b*) Each of the tests was correlated with English, mathematics, general science, and history individually, while (*c*) each of the elements of each test (31 in all) was correlated with all subjects combined, and with English, mathematics, general science and history individually. Correlations were also made between Army Alpha and university subjects of instruction for the first, second and third terms of 1 year.

[1] These four tests were given by S. R. Powers of the University of Minnesota to the pupils in the University of Arkansas Training High School in 1921.

In Tables I to X are the coefficients of correlation with their probable errors. In all 215 correlations were made in order to find out which group test or individual test correlated most highly with the subjects of instruction.

Of these intelligence tests Otis, Terman, and Army Alpha are familiar. The Miller test was developed by W. S. Miller of the University of Minnesota and is called "Test for High School Pupils." It was standardized by the administrative section of the high school conference. There are three parts to the test. The first consists of 40 disarranged sentences and requires 8 minutes to give. The second deals with causes and effects—*e.g.*, *fire* (hot house damage) "damage" being underlined correctly. There are 40 of these and 5 minutes are consumed in giving it. The third is the well known analogies and 6 minutes are given to answer the 40 analogies.

TABLE I.—CORRELATIONS OF FOUR GROUP TESTS OF INTELLIGENCE WITH THE COMBINED GRADE POINTS OF ALL SUBJECTS (67 CASES)

	FALL	WINTER	SPRING	AVERAGE
Otis	0.414 ± 0.066	0.527 ± 0.058	0.409 ± 0.067	0.450
Army Alpha	0.458 ± 0.064	0.449 ± 0.060	0.521 ± 0.060	0.476
Miller	0.460 ± 0.063	0.511 ± 0.059	0.457 ± 0.060	0.476
Terman	0.483 ± 0.062	0.476 ± 0.063	0.517 ± 0.059	0.492

Table I sets forth the correlations obtained for successive terms between each of the four group tests and the total number of grade points received throughout the year by the high school pupils in their several subjects of instruction. It will be noted that the Terman test stands highest in the fall quarter, Miller in the winter, and Army in

TABLE II.—CORRELATIONS BETWEEN EACH ELEMENT OF EACH OF THE GROUP TESTS OF INTELLIGENCE WITH THE COMBINED GRADE POINTS OF ALL SUBJECTS (67 CASES)

Otis	Subjects	Army	Subjects	Miller	Subjects	Terman	Subjects
1	0.269 ± 0.076	1	0.423 ± 0.067	1	0.466 ± 0.064	1	0.555 ± 0.057
2	0.342 ± 0.072	2	0.460 ± 0.064	2	0.380 ± 0.070	2	0.267 ± 0.076
3	0.305 ± 0.075	3	0.447 ± 0.065	3	0.441 ± 0.066	3	0.382 ± 0.070
4	0.323 ± 0.073	4	0.391 ± 0.069	..	.....	4	0.191 ± 0.079
5	0.479 ± 0.063	5	0.514 ± 0.059	..	.....	5	0.443 ± 0.066
6	0.374 ± 0.071	6	0.371 ± 0.070	..	.....	6	0.461 ± 0.063
7	0.414 ± 0.068	7	0.413 ± 0.068	..	.....	7	0.492 ± 0.062
8	0.293 ± 0.075	8	0.244 ± 0.077	..	.....	8	0.444 ± 0.066
9	0.298 ± 0.075	..	.....	..	.....	9	0.403 ± 0.069
10	0.136 ± 0.080	..	.....	..	.....	10	0.367 ± 0.071

the spring. Possibly the most significant finding here is the slight difference exhibited among the size of the correlations, no correlation being below 0.4 and no one being as high as 0.53. Moreover, no correlations are high, the results agreeing well with the 0.4 to 0.5 usually found by other investigators. Terman's tests average a little but not much above the others.

More interesting for our purposes is Table II which contains the correlations of each element of each group test with the total grade points of all subjects. Thus there are 31 correlation coefficients which vary in size from 0.136 for Otis-10, a test of memory, to 0.555 for Terman-1, a test of general information.

The highest correlations between individual tests and the total grade points of all subjects for the entire year were as follows:

Terman-1	General informaion	0.555 ± 0.057
Army -5	Mixed sentences	0.514 ± 0.059
Terman-7	Analogies	0.492 ± 0.062
Otis-5	Arithmetic problems	0.479 ± 0.063
Miller-1	Mixed sentences	0.466 ± 0.066

These correlations while not especially high are significant. It is to be noted that "mixed sentences" occurs twice. When these correlations are compared with those obtained with the four group tests and the total grades it is found that two of the individual tests mentioned in Tables I and II have higher correlations with all the grades than has the sum of the individual tests combined into the group of tests commonly known as the Otis tests of intelligence, Terman, etc. For example, the Otis test has a coefficient of 0.450; the Army test of 0.476; the Miller test of 0.476; and the Terman test of 0.492. It is evident, therefore, that Terman-1 (general information) which requires only 2 minutes to give and 1½ minutes to score would place pupils in groups, graded according to their capacities to learn as measured by the grades received, more correctly than would any one of the four group tests. Compare 2 minutes with an hour in giving, and 1½ minutes with 8 or 10 in scoring and the time for practical purposes seems clear.

The criticism might be raised, however, that a group of individuals might react differently when the test was given as a separate entity and for only a short period of time than they would when the test was one of a group extending over a longer period of time. I have no experimental evidence either for or against this problem. If this condition were found to be true it would not vitiate the use of this test

as one element of a group for prognostic purposes, nor prevent time being saved in scoring.

The correlations with English are seen in Table III. Here are found both the correlations obtained between the group tests as a whole and grades in English and the correlations between individual elements of each of the four group tests and English. The size of the correlations varies from Terman-2 with 0.14 to Terman-3 with 0.57.

TABLE III.—CORRELATIONS BETWEEN EACH ELEMENT OF EACH OF THE FOUR GROUP TESTS OF INTELLIGENCE AND GRADES IN ENGLISH (64 CASES)

Otis	English	Army	English	Miller	English	Terman	English
1	0.237 ± 0.077	1	0.301 ± 0.074	1	0.594 ± 0.053	1	0.433 ± 0.066
2	0.340 ± 0.071	2	0.386 ± 0.082	2	0.433 ± 0.066	2	0.141 ± 0.080
3	0.399 ± 0.069	3	0.544 ± 0.057	3	0.393 ± 0.069	3	0.572 ± 0.055
4	0.421 ± 0.067	4	0.527 ± 0.059	..	.....	4	0.386 ± 0.069
5	0.425 ± 0.067	5	0.409 ± 0.068	..	.....	5	0.253 ± 0.076
6	0.456 ± 0.065	6	0.279 ± 0.075	..	.....	6	0.492 ± 0.063
7	0.313 ± 0.073	7	0.301 ± 0.074	..	.....	7	0.552 ± 0.057
8	0.225 ± 0.078	8	0.388 ± 0.069	..	.....	8	0.459 ± 0.064
9	0.189 ± 0.074	..	.....	..	.....	9	0.383 ± 0.069
10	0.212 ± 0.078	..	.....	..	.....	10	0.387 ± 0.069
Total test....	0.466 ± 0.065	..	0.472 ± 0.065	..	0.564 ± 0.057	..	0.523 ± 0.061

English is one of the subjects with which a considerable number of individual tests correlate significantly, there being no fewer than 10 that have a correlation of 0.4 or above. The number of pupils was 64. These correlations are:

Miller-1	Mixed sentences	0.594 ± 0.053
Terman-3	Opposites	0.572 ± 0.055
Terman-7	Analogies	0.552 ± 0.057
Army-3	Reasons	0.544 ± 0.057
Army-4	Opposites	0.527 ± 0.059
Terman-6	Sentence meaning	0.492 ± 0.063
Terman-8	Mixed sentences	0.459 ± 0.064
Otis-6	Geometrical figures	0.456 ± 0.065
Otis-5	Arithmetic problems	0.425 ± 0.067
Otis-4	Proverbs	0.421 ± 0.067

These correlations between individual tests and English may be compared with the correlations with English when the group tests are used.

Otis	0.466 ± 0.065
Army	0.472 ± 0.065
Miller	0.564 ± 0.057
Terman	0.233 ± 0.061

One individual test, Miller-1 (mixed sentences), has a higher coefficient of correlation than the highest group test. It takes 8 minutes to give Miller-1 while it takes 20 minutes to give the Miller intelligence test; yet Miller-1 correlates with English more highly (0.594) than does the Miller intelligence test (0.564).

Again comparison may well be made between the correlations obtained for the whole year with those obtained for each term when the group tests were used. It is evident from Table III and Table IV that considered as a whole the correlations are somewhat smaller with term grades than with year grades. This is to be expected since, in general, to raise a correlation in the most convenient way is to lengthen the test.

TABLE IV.—CORRELATIONS BETWEEN FOUR GROUP TESTS OF INTELLIGENCE AND ENGLISH FOR SUCCESSIVE TERMS (67 CASES)

	FALL	WINTER	SPRING	AVERAGE
Otis	0.390 ± 0.067	0.405 ± 0.068	0.443 ± 0.067	0.412
Army	0.411 ± 0.067	0.401 ± 0.068	0.361 ± 0.072	0.391
Miller	0.426 ± 0.066	0.430 ± 0.067	0.430 ± 0.067	0.428
Terman	0.403 ± 0.067	0.364 ± 0.071	0.507 ± 0.062	0.424

The individual tests (Table V) which correlate most highly with the grades in mathematics are for the most part those concerned with mathematical operations. The one exception is the comparatively high correlation of hard oral directions. There are 47 cases.

TABLE V.—CORRELATION OBTAINED BETWEEN EACH ELEMENT OF THE FOUR GROUP TEST OF INTELLIGENCE AND GRADES IN MATHEMATICS (47 CASES)

Otis	Mathematics	Army	Mathematics	Miller	Mathematics	Terman	Mathematics
1	0.294 ± 0.090	1	0.608 ± 0.062	1	0.405 ± 0.082	1	0.316 ± 0.088
2	0.252 ± 0.092	2	0.521 ± 0.072	2	0.464 ± 0.077	2	0.207 ± 0.094
3	0.035 ± 0.098	3	0.406 ± 0.082	3	0.439 ± 0.073	3	0.291 ± 0.090
4	0.283 ± 0.090	4	0.233 ± 0.093	..	.....	4	0.074 ± 0.098
5	0.676 ± 0.053	5	0.212 ± 0.094	..	.....	5	0.529 ± 0.071
6	0.526 ± 0.071	6	0.471 ± 0.076	..	.....	6	0.249 ± 0.092
7	0.357 ± 0.086	7	0.454 ± 0.077	..	.....	7	0.342 ± 0.088
8	0.278 ± 0.090	8	0.274 ± 0.091	..	.....	8	0.300 ± 0.093
9	0.222 ± 0.093	..	.....	..	.....	9	0.202 ± 0.130
10	0.201 ± 0.094	..	.....	..	.....	10	0.448 ± 0.078
Group test...	0.430 ± 0.079	..	0.511 ± 0.073	..	0.456 ± 0.077	..	0.436 ± 0.079

The highest correlations are:

Otis-5	Arithmetic problems	0.676 ± 0.053
Army-1	Oral directions	0.608 ± 0.062
Terman-5	Arithmetic problems	0.529 ± 0.071
Otis-6	Geometric figures	0.526 ± 0.071
Army-2	Arithmetic problems	0.521 ± 0.071

It is to be noted that three of these are arithmetical problems and one is concerned with geometrical figures. Contrast these correlations with those obtained from the group tests.

Army	0.511 ± 0.073
Miller	0.456 ± 0.077
Otis	0.430 ± 0.079
Terman	0.436 ± 0.079

Therefore, if there are large classes in mathematics and divisions of them must be made, approximately homogeneous groups may be

TABLE VI.—CORRELATIONS BETWEEN FOUR GROUP TESTS OF INTELLIGENCE AND MATHEMATICS FOR SUCCESSIVE TERMS (47 CASES)

	FALL	WINTER	SPRING	AVERAGE
Otis	0.300 ± 0.080	0.374 ± 0.080	0.170 ± 0.090	0.281
Army	0.302 ± 0.079	0.366 ± 0.080	0.444 ± 0.085	0.371
Miller	0.299 ± 0.080	0.389 ± 0.080	0.338 ± 0.094	0.342
Terman	0.296 ± 0.080	0.287 ± 0.086	0.423 ± 0.086	0.335

obtained in a short period of time by using Otis-5. Furthermore, comparisons between the correlations obtained for the whole year in

TABLE VII.—CORRELATIONS OBTAINED BETWEEN EACH ELEMENT OF EACH OF THE FOUR GROUP TESTS OF INTELLIGENCE AND GRADES IN GENERAL SCIENCE (32 CASES)

Otis	General science	Army	General science	Miller	General science	Terman	General science
1	0.306 ± 0.108	1	0.491 ± 0.090	1	0.567 ± 0.098	1	0.502 ± 0.089
2	0.513 ± 0.088	2	0.513 ± 0.088	2	0.513 ± 0.086	2	0.323 ± 0.106
3	0.280 ± 0.110	3	0.468 ± 0.093	3	0.423 ± 0.080	3	0.468 ± 0.093
4	0.451 ± 0.095	4	0.514 ± 0.088	..	.....	4	0.402 ± 0.100
5	0.524 ± 0.086	5	0.548 ± 0.083	..	.....	5	0.381 ± 0.102
6	0.429 ± 0.097	6	0.245 ± 0.112	..	.....	6	0.516 ± 0.087
7	0.278 ± 0.110	7	0.392 ± 0.100	..	.....	7	0.546 ± 0.084
8	0.104 ± 0.118	8	0.427 ± 0.097	..	.....	8	0.413 ± 0.098
9	0.287 ± 0.109	..	....	..	.....	9	0.561 ± 0.082
10	0.345 ± 0.105	..	.....	..	.....	10	0.369 ± 0.103
Group test....	0.502 ± 0.89	..	0.596 ± 0.77	..	0.592 ± 0.78	..	0.636 ± 0.071

mathematics (Table V) and for those obtained for each term (Table VI) show smaller correlations for each term than for the whole year. In all cases save one all of the correlations for the whole year are above any one for the term.

General science considered from the standpoint of mental processes involved seems to be truly of a general nature and to resemble in its demand processes similar to those demanded of the individual tests (Table VII) since there are eleven of these tests with correlations higher than 0.5 with this subject. The number of cases was 32. These eleven highest correlations are:

Miller-1	Mixed sentences	0.567 ± 0.080
Terman-9	Classification	0.561 ± 0.080
Miller-5	Mixed sentences	0.548 ± 0.083
Terman-7	Analogies	0.546 ± 0.084
Miller-2	Giving effect of words	0.531 ± 0.086
Otis-5	Arithmetic problems	0.524 ± 0.086
Terman-6	Sentence meaning	0.516 ± 0.087
Army-4	Opposites	0.514 ± 0.088
Army-2	Arithmetic problems	0.513 ± 0.088
Otis-2	Opposites	0.513 ± 0.088
Terman-1	Information	0.502 ± 0.089

The group tests also correlate significantly with general science.

Otis	0.502 ± 0.089
Army	0.596 ± 0.077
Miller	0.592 ± 0.078
Terman	0.636 ± 0.071

In the case of general science three out of four of the group tests correlate more highly with general science than does any single one of the individual tests. Particularly noteworthy is the correlation of 0.636 in the case of the Terman test. Tables VII and VIII indicate as in the preceding subjects a smaller correlation for the term than for the three terms combined into 1 year.

TABLE VIII.—CORRELATIONS BETWEEN FOUR GROUP TESTS OF INTELLIGENCE AND GENERAL SCIENCE FOR SUCCESSIVE TERMS

	FALL	WINTER	SPRING	AVERAGE
Otis	0.432 ± 0.096	0.333 ± 0.103	0.456 ± 0.094	0.407
Army	0.539 ± 0.084	0.396 ± 0.098	0.506 ± 0.074	0.480
Miller	0.476 ± 0.092	0.388 ± 0.099	0.536 ± 0.080	0.466
Terman	0.485 ± 0.091	0.453 ± 0.094	0.586 ± 0.077	0.508

Since there are only 20 cases in history (Table IX) the results are not reliable. This unreliability is reflected in the unusually large probable errors.

TABLE IX.—CORRELATIONS OBTAINED BETWEEN EACH ELEMENT OF EACH OF FOUR GROUP TESTS OF INTELLIGENCE WITH THE GRADES IN HISTORY (20 CASES)

Otis	History	Army	History	Miller	History	Terman	History
1	−0.213 ± 0.149	1	0.253 ± 0.141	1	0.288 ± 0.138	1	0.362 ± 0.131
2	0.083 ± 0.149	2	0.115 ± 0.148	2	−0.047 ± 0.150	2	0.168 ± 0.146
3	0.193 ± 0.145	3	0.471 ± 0.117	3	0.033 ± 0.150	3	0.365 ± 0.127
4	0.247 ± 0.141	4	0.136 ± 0.147	..	.....	4	0.011 ± 0.151
5	0.079 ± 0.150	5	0.421 ± 0.124	..	.....	5	0.397 ± 0.127
6	0.001 ± 0.150	6	0.196 ± 0.145	..	.....	6	0.588 ± 0.098
7	0.199 ± 0.147	7	0.007 ± 0.151	..	.....	7	0.364 ± 0.131
8	0.135 ± 0.148	8	0.237 ± 0.142	..	.....	8	0.364 ± 0.131
9	0.392 ± 0.127	..	.....	..	.....	9	0.429 ± 0.121
10	−0.008 ± 0.150	..	.....	..	.....	10	0.117 ± 0.149
Group test....	0.262 ± 0.140	..	0.319 ± 0.136	..	0.168 ± 0.148	..	0.408 ± 0.121

The following individual tests correlated highest with history:

Terman-6	Sentence meaning	0.588 ± 0.098
Army-3	Selecting reasons	0.471 ± 0.117
Terman-9	Classification	0.429 ± 0.121
Army-5	Mixed sentences	0.421 ± 0.124

The ability to get meaning from a series of sentences some of which are false has the highest correlation with history. The group tests which in no case correlate so highly with history have the following coefficients:

Army	0.319 ± 0.136
Terman	0.408 ± 0.121
Otis	0.262 ± 0.140
Miller	0.168 ± 0.148

By reference to Table X it is once more evident that the correlations for 1 year are higher in general than those for individual terms.

TABLE X.—CORRELATIONS BETWEEN FOUR GROUP TESTS OF INTELLIGENCE AND HISTORY FOR SUCCESSIVE TERMS (20 CASES)

	Fall	Winter	Spring	Average
Otis	0.248 ± 0.146	0.319 ± 0.120	0.453 ± 0.114	0.340
Army	0.341 ± 0.130	0.256 ± 0.120	0.258 ± 0.132	0.285
Miller	0.198 ± 0.141	0.101 ± 0.135	0.141 ± 0.132	0.147
Terman	0.419 ± 0.124	0.313 ± 0.125	0.335 ± 0.122	0.356

The correlations of Army Alpha, Otis and Miller are unreliable because they are not three times as large as their respective probable

errors, or, if the more stringent test of the coefficient's being four times its probable error should be applied, not one of the group tests and only Terman-6 (sentence meaning) and Army-3 (selecting reasons) are sufficiently reliable. Thus again the advantages of the individual tests are apparent.

In general except in the case of general science it is not at all difficult to find individual tests which correlate more highly with any or all subjects of instruction than the widely used group tests. When the value of each element of each test is accurately determined the question of the purpose to which a test is to be put will, of necessity, be an important item in its selection. The particular question raised here is one of prognosis, the human engineer wishing to know how prophetic the standing in a mental test is of the standing in some life situation.

In this paper no attempt has been made to compare the findings with those obtained in other studies. This, if complete, would mean a paper two or three times as long as the present one. However, the correlations obtained with the Army Alpha and grades of university freshmen might be set forth for comparative purposes:[1] From Table XI it may be seen that, in general the coefficients of correlation are somewhat but not much higher when the high school grades are used than when the university grades are used.

TABLE XI.—CORRELATIONS OF ARMY ALPHA AND GRADES RECEIVED IN THE FRESHMAN YEAR OF THE UNIVERSITY

	First term	Second term	Third term	Cases
Alpha and average grades	0.485	0.28	0.278	304
Alpha and English	0.517	0.42	0.29	265
Alpha and mathematics	0.213	0.47	0.02	94
Alpha and language	0.313	0.28	0.32	93
Alpha and history	0.540	0.07	0.09	37
Alpha and science	0.448	0.52	0.43	247
Average	0.419	0.34	0.234	

The next problem is that of discovering which one of the group tests is best for prophecying standings in school subjects. What is said is based on simple correlations. There might be combinations of tests within each group which would correlate more highly than any

[1] Jordan, A. M.: Some Results and Correlations of Army Alpha Tests. *School and Society*, March 20, 1920.

of the correlations mentioned below. That group test is defined as "best" from which may be obtained either from the group as a whole or from any individual test the highest correlations in the subject considered.

TABLE XII.—ARRANGEMENT OF CORRELATIONS TO SHOW HIGHEST CORRELATIONS OF EACH GROUP TEST WITH ALL SUBJECTS COMBINED AND WITH ENGLISH, MATHEMATICS, GENERAL SCIENCE, AND HISTORY

	All subjects	English	Mathematics	General science	History
Otis	5.....0.479 Group...0.450	4.....0.421 5.....0.425 6.....0.456 Group...0.466	5.....0.676 6.....0.526 Group...0.430	2.....0.513 5.....0.524 Group...0.502	Group....0.262
Army	5.....0.514 Group...0.476	3.....0.544 4.....0.527 Group...0.472	1.....0.608 2.....0.521 Group...0.511	2.....0.513 4.....0.514 5.....0.548 Group...0.456	3.....0.471 5.....0.421 Group...0.319
Miller	1.....0.466 Group...0.476	1.....0.594 Group...0.456	..... Group...0.456	1.....0.567 2.....0.531 Group...0.592	Group...0.168
Terman	1.....0.555 7.....0.492 Group...0.492	3.....0.572 7.....0.552 6.....0.492 8.....0.459 Group...0.436	5.....0.529 .... Group...0.466	1.....0.502 6.....0.516 7.....0.546 9 0.561 Group...0.636	6.....0.588 9.....0.429 Group...0.408

In Table XII, "group" means all the elements of the test combined; the small numbers before the coefficients of correlation refer to the elements of the group test; thus, in the first line one would read "Otis, element five (arithmetical reasoning) has a correlation with all subjects of 0.479."

The inferences from Table XI are as follows:

1. For all subjects combined *Terman* stands above the rest because of Test 1 with a coefficient of 0.555, and because the correlation between the group of tests and all subjects is 0.492.

2. For English, *Miller* is best because by using Test I alone one gets a coefficient of correlation of 0.594.

3. For mathematics *Otis* is best since by using Test V alone a correlation of 0.676 may be obtained.

4. For general science, *Terman* is best since this group of tests gives an r of 0.636.

5. For history, *Terman* is best since Test VI gives a correlation of 0.588.

By using Terman's tests of intelligence, therefore, one gets the best test for general prognosis (0.55), best for general science (0.64), best for history (0.59), the second best for English (0.572) the third best for mathematics (0.529). The question of the value of these tests for general intelligence is not raised but only the question of prognosis for school work.

Discussion and Conclusion

Some tests are better for certain purposes than are others but in no case has a really high correlation been found. The highest correlation obtained is 0.68. I have calculated from a table of Thorndike's which is based on tenths that if we used fifths instead of tenths and wish to prophecy how many cases in each fifth of one series would be in the corresponding fifth in the related series, this number is 32 in 100 when the correlation is 0.5; the corresponding numbers when thirds are used are approximately 46 out of 100 with a correlation of 0.5. The figures for a correlation of 0.7 are, for fifths, 40 out of a 100; for thirds, 55 out of a 100; for a correlation of 0.9, 57 out of 100 for fifths, and 71 out of 100 for thirds. Thus it is seen that for reliable prognostic purposes we would need a correlation around 0.9. Which is manifestly impossible to obtain with grades since the self correlation of grades is not this high. These facts point to the distance we are from tests that will accurately prophecy. It may be that we shall be compelled to turn to the tests of specific abilities. The unexpected size of the correlations of some elements of these so-called general intelligence tests with grades also points as far as grades are concerned to the need of tests prognostic of standing in specific subjects such as the Rogers diagnostic tests for mathematics.

LANGUAGE ERROR TESTS[1]

(Concluded from the September issue)

G. M. WILSON

Boston University

The next problem was to bring these errors into new stories in satisfactory form and then re-test them to see whether or not there would result three compositions of approximately equal values. Again an advanced class was used in writing stories in an attempt to make use of these errors in satisfactory form for testing. With these as a basis, three tests were finally developed—Test A, "Saturday Morning;" Test B, "A Fishing Trip;" Test C, "An Accident." Through the cooperation of the superintendents of schools in Sioux City, Iowa, and Duluth, Minnesota, it was possible to try out these tests.

The combined medians for Sioux City and Duluth, based upon a total of 6965 tests, are shown in Table VII.

TABLE VII.—MEDIANS FOR SIOUX CITY AND DULUTH (COMBINED)

Grade	III	IV	V	VI	VII	VIII	IX	X	XI	XII
Test A.....	10	17	20	22	23	23	25	25	25	26
Test B.....	15	13	19	20	21	22	25	26	27	26
Test C.....	17	15	20	23	22	24	25	26	26	26

The nature of the tests and their equality one with the other will be much more evident, however, from a showing of the total distributions. These are indicated herewith in Tables VIII, IX, and X. It is evident from these tables that Tests A, B, and C approach equality.

[1] The "Wilson Language Error Tests," will be published by the World Book Company, Yonkers, N. Y.

TABLE VIII.—DISTRIBUTION OF SCORES FOR RIGHTS
Test A. "Saturday Morning"
(Duluth and Sioux City combined)

Score	Grade										
	III	IV	V	VI	VII	VIII	IX	X	XI	XII	
0	1										
1	6	1									
2	4										
3	11	1									
4	10	1									
5	17	1									
6	21	3									
7	19	2	1								
8	15	4	3	1							
9	11	4									
10	17	6	2	...	1						
11	9	12	2	...	...	1					
12	14	13	5	1	2						
13	13	10	6	2	1						
14	11	13	5	5	1	1					
15	12	18	8	8	4	...	1				
16	11	15	12	3	4	1	...	...	1		
17	13	18	7	9	10	5					
18	8	11	14	14	13	6	2	1			
19	8	20	15	18	16	15	...	3	1		
20	5	15	20	20	24	20	2	4	1	1	
21	7	21	24	26	29	29	2	3	0	2	
22	1	6	31	36	23	37	4	12	6	2	
23	3	14	26	34	32	37	13	8	8	9	
24	...	6	16	49	43	47	8	16	8	8	
25	...	6	10	21	39	45	14	23	9	12	
26	2	1	8	8	23	33	15	31	13	18	
27	...	...	2	3	10	12	11	22	14	25	
28	...	...	3	...	4	5	...	4	5	6	
Total.....	248	222	220	258	279	294	72	127	66	83	1869
Q. 1.....	6	13	18	20	20	21	23	24	23	25	
Median.....	10	17	20	22	23	23	25	25	25	26	
Q. 3.....	15	20	23	24	25	25	26	26	27	27	

TABLE IX.—DISTRIBUTION OF SCORES FOR RIGHTS
Test B. "A Fishing Trip"

Score	Grade										
	III	IV	V	VI	VII	VIII	IX	X	XI	XII	
0	...	4									
1	...	4	1								
2	2	6									
3	1	3									
4	...	9	2								
5	1	8	2	1							
6	1	11	2	...	3						
7	6	5	1	2	3	...	1				
8	6	8	4	4	6						
9	13	5	6	2	5						
10	6	14	5	7	7						
11	14	12	8	3	3	1	2				
12	7	5	10	3	13	1	1				
13	20	10	5	8	12	1					
14	14	14	3	7	5	11					
15	16	5	12	10	19	7	1				
16	15	11	10	6	24	9	2				
17	16	9	17	8	29	14	1	...	1		
18	14	9	12	10	28	20	6	1			
19	9	10	12	8	31	24	5	...	..	1	
20	14	8	10	13	36	30	5	1	..	2	
21	6	9	15	8	39	40	11	2	3	3	
22	10	5	17	20	36	38	17	5	3	2	
23	5	6	10	23	36	29	19	10	3	7	
24	7	6	9	12	40	31	24	19	7	7	
25	4	3	14	8	33	30	36	16	8	16	
26	1	2	10	9	11	20	27	23	14	15	
27	...	5	4	8	22	14	27	26	23	10	
28	...	...	1	1	3	4	8	8	21	22	
29	...	...	2	...	3	4	8	4	6	4	
Total.....	208	206	203	181	447	328	201	115	89	89	2067
Q. 1.....	12	8	15	15	17	19	21	24	25	25	
Median.....	15	13	19	20	20	22	25	26	27	26	
Q. 3.....	19	18	22	22	24	24	28	27	28	28	

TABLE X.—DISTRIBUTION OF SCORES FOR RIGHTS
Test C. "An Accident"
(Duluth and Sioux City combined)

Score	Grade										
	III	IV	V	VI	VII	VIII	IX	X	XI	XII	
0											
1	2	...	1								
2	2	1									
3	1	4									
4	2	2	1	1							
5	4	5									
6	2	2									
7	1	6	3	1							
8	9	13	4	1							
9	4	6	4	1							
10	5	11	8	1	1	1					
11	6	16	10	...	1	2					
12	8	18	5	2	2	1					
13	8	10	4	5	2						
14	10	14	13	7	7	1	1				
15	10	14	7	4	2	3					
16	8	15	17	14	13	9	2				
17	10	17	15	11	15	7	2	2			
18	10	16	11	11	22	11	3	1			
19	10	9	19	17	20	19	5	1			
20	10	11	24	17	22	17	13	3	2	1	
21	16	10	10	17	27	26	10	5	4		
22	7	12	19	21	31	25	4	11	6		
23	7	9	24	34	49	35	19	16	8		
24	12	11	14	20	35	42	36	19	10		
25	9	4	23	27	34	36	31	16	17	2	
26	5	2	15	27	23	49	28	35	20	2	
27	1	1	15	15	12	29	36	27	26	2	
28	1	...	15	8	6	18	14	10	19		
29	...	...	5	9	1	9	2	10	7		
Total.....	180	239	286	271	325	340	206	156	119	7	2129
Q. 1.....	12	11	16	19	19	21	23	23	24	25	
Median.....	17	16	20	23	22	24	25	26	26	26	
Q. 3.....	21	20	25	25	24	26	27	27	27	27	

Further Evidences of Equality.—The limits of this article will not permit submitting extended data under this heading, but several things were done. All pointed strongly to the conclusion that the three tests are practically of equal value. The percentage of pupils in Grades III, VI, and X making each error in each test was figured. The median percentages for the different grades run surprisingly close together.

The 12-year-old group was separated and distributed. In all three tests, A, B and C, the median for the 12 year olds was 23 and the quartiles were either on the same point or only one step removed.

Tests B and C were given at the same time to pupils in grades III to VIII, inclusive, in a small city system. On the basis of 103 pairs thus secured, the coefficient of correlation was figured. The results showed a positive coefficient of 0.901.

On the basis of the above evidence, it seems safe to conclude that the three tests are, for all practical purposes, of equal value and may be used interchangeably. There are 28 errors in Test A and 29 in each of tests B and C, but the score for rights is practically the same for any of the three tests.

Probable Error of a Score.—Under the direction of Dr. Arthur S. Otis, the probable error of a score by the Difference Method[1] was figured. The resulting probable error was 2.24. This shows a high degree of reliability for the test. It confirms the high coefficient of correlation.

The Tests.—Tests A, B, and C in the form in which they were used follow herewith. It is planned later to revise and simplify the directions leaving the tests themselves in their present form:

TEST A.—CORRECTING LANGUAGE ERRORS. (*A Game*)

Name.....Grade.....Age.....

Town.....School.....Date.....

Directions for the Game.—(To be read by the teacher, the pupils following.) This is a little *game* in which the pupil plays teacher, and corrects a composition written by a pupil. Correct by drawing a single line through words or expressions used incorrectly, and placing the correct words above them. For example, if you had the following sentence to correct: "He has went home" you would correct it by drawing a single line through *went* and writing *gone* above it. Make all changes necessary to secure correctness. Work at your usual rate. You will be given reasonable time in which to complete your work. When you have finished, turn

[1] Otis, Arthur S: Reliability of Binet Scale and Pedagogical Scales. *J. of Educational Research*, September, 1921.

the sheet right side down and leave it on your desk. All will be permitted to finish the work unless too slow.

The composition which you are to correct follows herewith:

SATURDAY MORNING

Saturday morning is a busy time to are house. A feller has a good chance to work. Me and Dorothy we divide the tasks between us. Then we race to see who will finish first. Last Saturday I taken the breakfast dishes as one of my tasks. I am especial fond of washing dishes. You should have saw me work. I wanted to get through so as I could play. John he called up at eleven o'clock to see if I might play with him. I had two rooms to dust before I could go. John seen that I wouldn't leave my work until I had did all of it. He brought over some doughnuts and give them to me. I sure appreciated them doughnuts. Then John helped me. When we had finished, I suggested playing marbles until time for dinner. "I ain't got any marbles," said John. "They comes mighty handy," I replied. Then I give him some of mine. I had got to many for my bag. I and John enjoy playing marbles. When dinner was ready, mother invited John to stay. "If I was sure my mother wouldn't care, I would like to stay," he replied. John he seen that he was really wanted so he telephoned his mother. He enjoyed the dinner and ate heartily. When them apples were passed, John wanted one, but he couldn't eat no more. After dinner we had another game of marbles. I hope John may come over again.

TEST B.—CORRECTING LANGUAGE ERRORS. (*A Game*)

Name.....Grade.....Age.....

Town.....School.....Date.....

Directions for the Game.—(To be read by the teacher, the pupils following.) This is a little *game* in which the pupil plays teacher, and corrects a composition written by a pupil. Correct by drawing a single line through words or expressions used incorrectly, and placing the correct words above them. For example, if you had the following sentence to correct: "He has went home" you would correct it by drawing a single line through *went* and writing *gone* above it. Make all changes necessary to secure correctness. Work at your usual rate. You will be given reasonable time in which to complete your work. When you have finished, turn the sheet right side down and leave it on your desk. All will be permitted to finish the work unless too slow.

The composition which you are to correct follows herewith:

A FISHING TRIP

John he is awful good to me. He once't ask me to go fishing and said that he could learn me to be a good fisherman on no time. He

had saw some men git a great many fish out of a deep hole about a mile up the river. He said that he had watched them until they got tired fishing. He seen them leave with a large sack full. I agreed to go with him. "We hain't got any bamboo poles," he said, "the folks haven't none left over from last year. Good poles is difficult to find." John give me the lunch to carry. We et our lunch before we done any fishing. I sit the table while John cut two poles and fastened the lines to them. He baited my hook hisself and told me to throw it in. I felt a bite at once and jerked the line. It was a large catfish. I was afraid the line would break. John said that the line was made good and had held to many big fish to break easily. I landed the fish but we didn't catch any more. We wanted to git another one so that each of us would have a fish to take home. We started and went home early. John said, "You can't never tell about beginners luck." John and I is good friends. As I neared home, I seen my little brother coming down the street. He had came to meet us. "I have got a big one," said I, as I showed him the fish. The fish was the main part of are supper that evening.

TEST C.—CORRECTING LANGUAGE ERRORS. (*A Game*)

Name.....Grade.....Age.....
Town.....School.....Date.....

Directions for the Game.—(To be read by the teacher, the pupils following.) This is a little *game* in which the pupil plays teacher, and corrects a composition written by a pupil. Correct by drawing a single line through words or expressions used incorrectly, and placing the correct words above them. For example, if you had the following sentence to correct: "He has went home" you would correct it by drawing a single line through *went* and writing *gone* above it. Make all changes necessary to secure correctness. Work at your usual rate. You will be given reasonable time in which to complete your work. When you have finished, turn the sheet right side down and leave it on your desk. All will be permitted to finish the work unless too slow.

The composition which you are to correct follows herewith:

AN ACCIDENT

One Friday afternoon are teacher she asked us if we wanted to go to the woods. It was an awful nice day! Ain't it fun to play in the woods on such days? Their was a woods near the school house. There was lots of flowers in bloom. John and me wanted to pick them flowers so's we could take some home to mother, but the others did not want to wait for us. We had not went far when we saw too squirrels. They run away from us. John ain't never seen such funny

little animals. He asked if he might throw a stone at them. He done it and the stone bounded back and struck him on the head. He had to pay up for it because their was a large bump on his head. We hadn't no medicine with us, so we had to go home to git some for him. If I were him, I would let the squirrels play next time. Me and William felt sorry for him. His mother give us some apples for bringing him home. There orchard was full of apples. They never had as many before. The apples was picked and lay in great piles under the trees. The apples which we received helped to make up for the disappointment in having to come home early.

It is hoped that superintendents and measurement experts throughout the country may make use of these language error tests, sending summary of the returns to the writer. Actually working with the tests and noting what can be done with them to determine the language ability of school children and the points on which a particular class is weak, will carry conviction that the tests are of real value.[1]

---

[1] The writer has been impressed as he has been using these language error tests that they have exceptionally high diagnostic value, and he hopes to set this forth in another article, showing at the same time just how to use and score the tests for diagnostic purposes.

THE RELATION OF RHYTHM TO THE HAND-WRITING MOVEMENT

An Experimental Study

PAUL V. WEST

Assistant Prof. of Education, Univ. of Wisconsin

Much general interest has been attached in recent years to the problems relating to temporal controls of the handwriting movement. Chief among these problems are those having to do with rhythm. What is meant by the term "rhythmical writing?" Do better writers perform more rhythmically than poorer writers? What is the effect of imposed rhythm on performance? Is arm or finger movement better adapted to rhythmitization? What are some of the most pertinent suggestions governing the use of and the emphasis on rhythm in penmanship instruction?

Psychological experiments in rhythm have not been without significance and value in their general bearing on these inquiries, but direct experimentation in the field of handwriting is essential to satisfactory solutions. The investigations of Freeman[1] and Nutt[2] are outstanding reports which have contributed substantially to a better knowledge of rhythmic movement in handwriting. In brief they show that temporal rhythm in writing increases with age of the writer and is assisted by increase of speed. They also note that it is not correlated with good form, in fact has a tendency to interfere with quality in certain cases, and that there is little distinction, if any, in finger and arm movement as to influence on rhythmitization. These investigators have arbitrarily defined rhythmic movement as the tendency to give succeeding strokes the same time emphasis—in other words, rhythm is regarded as a simple uniformity in duration of strokes.

Two types of experiment here noted were undertaken with the desire primarily to gain more specific information regarding rhythmic movement as related to the penmanship of good and poor writers of both adult and child groups. The first of these was carried out by having the subject make to and fro movements, as in writing, with finger and arm movement, without and with imposed rhythm. By a system of electric contacts a kymographic record was made which

[1] Freeman, F. N.: The Handwriting Movement, *Educ. Mon.*, 1918, Vol. II, 2, 3.

[2] Nutt, H. W.: Rhythm in Handwriting, *El. Sc. Jour.*, Feb., 1917.

showed the relative amount of time spent on the stroke and the rest at the terminus of the stroke. The second experiment was effected by a photographic method by means of which the actual writing process was analyzed in terms of distance covered during each fiftieth of a second. This method proved very effective and accurate.

The first experiment did not provide means of studying actual writing movement as did the latter, but did reveal characteristic behavior of the respective sets of muscles used in handwriting. The criteria of rhythm here considered were (1) the approach to temporal uniformity of successive total strokes, from the beginning of each stroke to the beginning of its successor, and (2) equality of time spent on similar stroke elements (the period of movement or stroke proper, and the period of rest at the end of the movement) as well as their combination in the total stroke unit.

The results were analyzed so as to show the percentage of variation within these various units as summarized in Table I. The data of this table are presented graphically in Diagram I so as to make clear the effect of the imposed rhythm. The variation of the younger children is generally greater than that of any of the other groups.

TABLE 1.—AVERAGE PER CENT OF VARIABILITY FROM TEMPORAL UNIFORMITY

Group		Rest		Stroke		Total		
		Finger	Arm	Finger	Arm	Finger	Arm	
Adult	Good	13.7	18.2	10.5	8.7	6.8	6.4	Spontaneous
	Poor	15.4	13.6	11.0	7.9	6.3	6.5	
	Average	14.5	15.9	10.7	8.3	6.5	6.4	
Adult	Good	14.3	16.8	10.8	9.7	8.0	7.1	Imposed
	Poor	16.1	17.3	13.0	11.0	8.4	7.8	
	Average	15.2	17.0	11.9	10.3	8.2	7.4	
Child	Young	21.8	20.0	24.1	14.4	13.7	8.8	Spontaneous
	Older	16.3	15.8	13.0	9.4	14.4	7.9	
	Average	19.0	17.9	18.5	11.9	14.0	8.3	
Child	Young	18.6	22.9	20.5	17.2	11.8	12.3	Imposed
	Older	15.8	21.7	15.5	14.9	8.8	9.7	
	Average	17.2	22.3	18.0	16.0	10.3	11.0	

Although there are many cases of overlapping the better adult writers maintain a slightly superior record of regularity. The rhythmic reaction to the total period as a unit is quite noticeable, for all groups here show less relative variation than in either the stroke or rest.

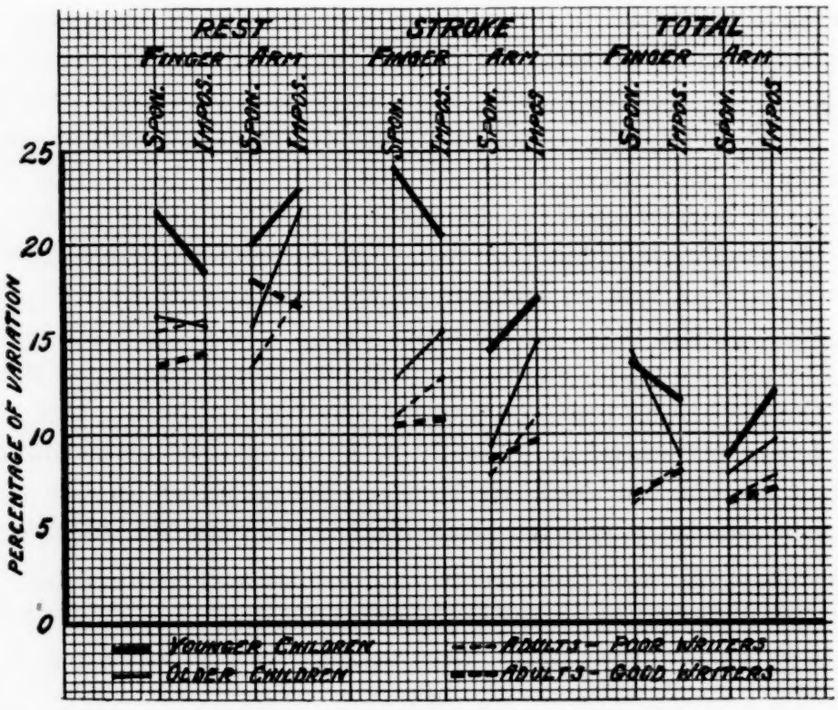


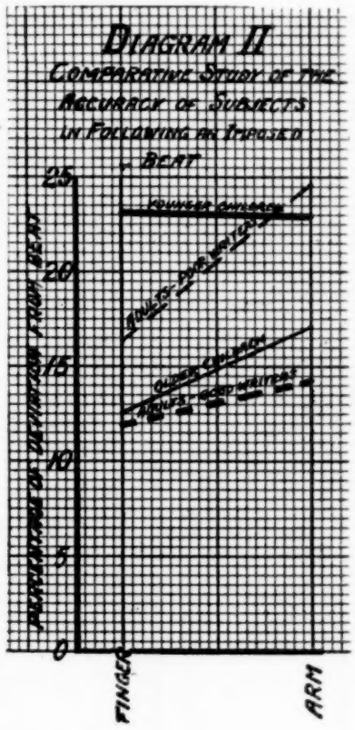
DIAGRAM I.—AVERAGE PER CENT OF VARIABILITY FROM TEMPORAL UNIFORMITY (Showing comparison of good and poor adult writers and younger and older children, under conditions of spontaneous and imposed rhythm, with the use of finger and arm movement.)

The finger and arm movements are doubtless distinctively significant for rhythmic behavior. The arm, because of its greater mass weight and consequent inertia, seems to be more easily adapted to natural rhythmic movement than the finger, especially in the case of children who have not yet gotten as ready control of the less gross finger mechanism. On the other hand, when the child attempts to follow a set rhythm, the inertia of the arm is manifest in the fact that the arm is less subject to rhythmitization than the fingers. Rhythmic habituation is apparently attained in arm movement, at least to marked extent, while the fingers, being unhabituated, yield more readily to the imposed beat.

This factor of habituation is marked also in comparison of the various groups as to the effect of an imposed rhythm on variation.

For adults, who have already acquired a comparatively high degree of rhythmic movement, probably as a more or less fixed norm of performance, the amount of variation is greater in both finger and arm movement when the rhythmic beat is followed than when the action is natural. The better writers show greater power of adaptation than the poor writers in this respect, however.

An investigation of the accuracy with which the imposed beat is followed reveals the fact that the poorer adult writers are less accurate than the older children, while the better writers show comparatively high ability to follow the rhythmic beat closely (Diagram II). In general the better writers and older children are able to anticipate the beat while the poorer writers and young children show a distinct tendency to lag behind the beat. The arm is found to be less accurate than the fingers in attaining the rhythmic pace set, except in the case of the younger children, who have not yet gained control of the finger muscles to the same degree as those who are older.



The imposed rhythm is found to increase the proportionate amount of time spent at rest for all groups, but this is especially notable in the case of the younger children, and when finger movement is used. The total duration of the stroke is increased, but the time spent on the stroke proper tends to remain constant.

In the second experiment the handwriting movement of 20 different subjects was investigated in a great deal of detail. Seven of the subjects were children ranging in ages from 7 to 16 years. Six of the remaining number, who were adults, were poor writers and seven were good writers as judged by the degree of evident coordination in their writing. Each subject wrote an indirect running oval and the majority wrote a sentence so that a study might be made of a simple repetitive form as well as a sample of characteristic penmanship.

In Figure I is given a reproduction of the written record of the 7-year-old child in making a running oval, with the successive 50th-second intervals indicated by the spaces between the cross bars.

The irregularity in form and in progression is evident. The poor organization of speed of movement is made very clear in Diagram III where the speed curve is in part compared with the curve of a record for the same exercise as made by a good adult writer. In the former there is no even progression of speed, and no adaptation of speed to the form to be produced, as in the latter. Also the many changes of speed are almost beyond comprehension and offer quite a contrast to the regular rise and fall of speed at the middle and end of strokes respectively which characterize the adult record.

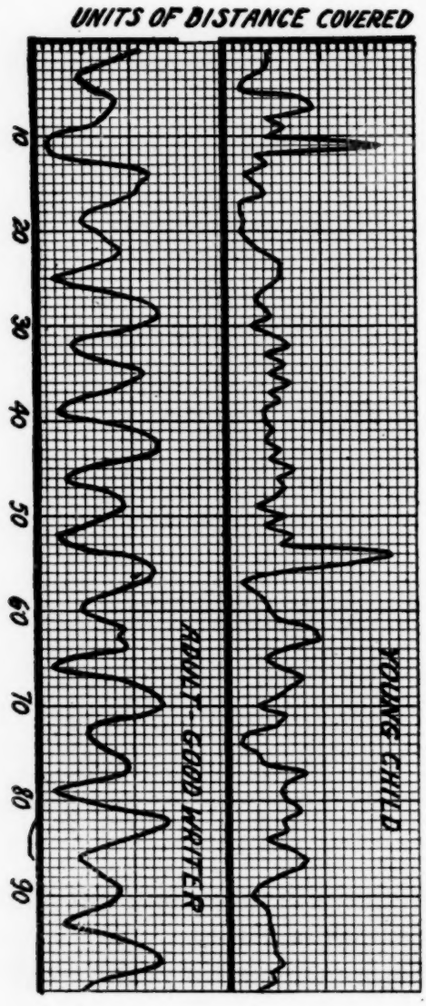


DIAGRAM III.—COMPARISON OF SPEED CURVES OF THE YOUNG CHILD AND GOOD ADULT WRITER IN WRITING THE RUNNING OVAL.

Speed curves as thus constructed yield readily to rhythmic analysis since the number of units of time spent on each stroke can be accurately computed. An analysis of the records for the repetitive forms showed that the good writers among the adults made less variation from their average length of time both on the up stroke and down stroke as well as on the total of both than the poor writers, but very poor writers may show better natural rhythm than most of the good writers. The children, although in some cases showing a high degree of rhythm, are on the whole comparatively low in this respect.

When the records made in the writing of words were analyzed the rhythm was found to be very much lower, due undoubtedly to the irregularity in the length of strokes being constructed, as well as their complexity. While the individuals tend to give the same average length of time to strokes in the two writings of the same word or different words, the tendency is to devote unequal amounts of time to the respective strokes within a word. In general, the longer strokes receive the greater time emphasis, although this proportion is modified by the complexity of the stroke.

Ovals were analyzed to see what was the effect of curvature on the speed attained within the stroke. In Figure II is shown a portion of an oval broken up for measurement into seven arced segments having their centers of radiation in order as indicated by the numerals 1 to 7. It is notable that the radial lengths differ greatly. In the complete form of this particular oval the correlation between radial length and

FIGURE ONE
Writing of Seven Year Old Child Showing Comparative Distances Covered in each 50th of a Second.

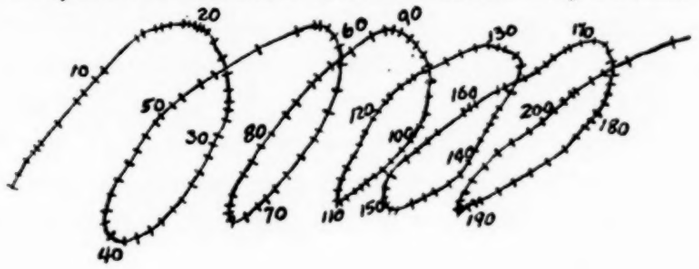
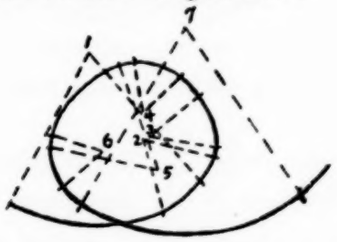


FIGURE TWO
Showing Method of Analyzing Curve into Arced Segments.



maximum speed within the arc for 33 segments was found to be very high by the Spearman Rank Method, the coefficient being 0.96. The line of regression however indicates a curved rather than a straight line relationship with a falling off in the regular increment of speed after a certain radial length has been reached, and a tendency for the rate to reach a fixed limit as the arc approaches a straight line form. This relation was found to be quite generally characteristic of the subjects. The poorer writers and children show very much lower correlation between the maximum speed and curvature of the arc than do the good writers and also exhibit a more marked tendency to slow down the speed relative to the increase of the radial length of the arc.

If, as the strokes decrease in curvature, the maximum speed within the stroke were to remain constant, the time must be the variable, varying directly with the radial length of the stroke in case the stroke

resembles a simple arc. But were the speed to vary in exact proportion as the radial length the time would remain constant. The closer approach to this constancy on the part of the good writers indicates a marked tendency toward temporal rhythm. If the strokes used in writing could be so made as to be perfectly arced segments, it is possible that they might be constructed in about the same length of time regardless of size or degree of curvature. But the form of the written word does not ordinarily yield to such simple temporal analysis.

Experimentation with a few subjects in the following of a rhythmic beat while writing words showed that a writer with habituated speed and rhythm would be slowed down by any beat that was slow enough to be consciously followed. The regularity of time thus induced in the successive strokes was also in marked contrast to the normal behavior of the subject, or of any one else examined, and resulted in uncertainties of speed progress which affected harmfully the quality of the written form.

The fluency, freedom and smoothness of movement which characterizes the good writer generally is the fundamental "rhythm" of handwriting. It is due to the unhampered interaction and coordination of the fingers and arm as free swinging levers, and with a ready control of the motor elements at every instant, so there is a definite relation between the form being constructed and the speed and time relationships.

The use of the rhythmic guide in penmanship instruction must be carefully supervised and scientificaly directed. The value of such a guide in the early years is unquestioned, as an aid in organizing the writing into unit strokes, and also as an aid in facile movement. For purposes of temporal rhythmitization of movement an imposed rhythm is not thoroughly adaptable to the process of constructing word forms, but may be used to better advantage with simple repetitive forms. After the subject has developed habits of natural rhythm the use of the rhythmic guide may result in a retardation of speed and a detrimental effect on form. Little is known as to the extent to which an imposed rhythm may be acquired as a fixed habit by practice. It appears that its chief value must lie in the development of a smoothly regulated type of movement in which the writer has a controlled adjustment of speed to form. The training process must involve due consideration of all the muscles naturally used in handwriting, as well as all individual differences due to age and natural tendencies to rhythmic movement.

NEW PUBLICATIONS IN EDUCATIONAL PSYCHOLOGY AND RELATED FIELDS OF EDUCATION

New Studies in Reading

A Consideration of Essential Principles of Teaching Reading and Literature.—Professor Leonard's new book[1] on reading is an exceedingly stimulating and constructive volume embodying a point of view which issues from a sound conception of educational values. Mr. Leonard has drawn heavily on his experiences to illustrate and clarify his discussion and where that did not suffice he quotes generously from the experience of numerous co-workers. The abundance and quality of these "samples" bridge the gap so often left between theory and practice and makes the book as readable as it is instructive. The discussion of tests and measurements are illustrated by numerous well chosen graphs. Illustrations depict actual school dramatizations and library conditions. One appendix contains a bibliography related to each chapter of the book; another sixty-three page appendix of annotated booklists for children is arranged by grades as well as by subjects and covers all grades. The book is very thoroughly indexed.

L. Z.

---

3. *A Critical Study of the Content of the School Reading Course.*—Critics of the content of the elementary school curriculum in reading will appreciate the outcomes of Dr. Uhl's study[2] based on the reactions of over 2000 teachers in more than 100 cities and on an analysis of over 500 pupil scores and judgments. These were used in constructing standards for rating reading selections[3] proposed for use in the selection, elimination and grade placement of materials.

The method of the investigation and the treatment of data lead

---

[1] Leonard, Sterling A.: "Essential Principles of Teaching Reading and Literature." J. B. Lippincott Co., Philadelphia, 1922, p. 460.

[2] Uhl, Willis L.: Scientific Determination of the Content of the Elementary School Course in Reading. "University of Wisconsin Studies in the Social Sciences and History," No. 4., Madison, 1921, p. 152.

[3] Rating Scales for Reading Selections.

to significant findings which are clearly set forth in tabular and graphical form. The representation of changes in pupils' interest in and comprehension of certain selections provides a new approach to grade placement and should be reflected in a happier and wiser choice of literary matter for use in the grades.

Interest and difficulty are considered separately in the rating scales with which the study concludes, and each value is illustrated by several well known samples.

L. Z.

---

4. *Reading Research Interpreted for Teachers.*—Dr. C. T. Gray has sought to restate the results of his own investigations and others in a volume designed for use in special methods courses and reading circles.[1] In no case does he dogmatize without presenting scientific evidence. Reading ability is analyzed from four standpoints and the possibilities of diagnosis are pointed out from the standpoint of standard tests, perception, motor elements and the higher mental activities of reading. This analysis and synthetic summary comprise the 14 chapters of Part I.

Part II discusses the methods and materials necessary for diagnostic observation and measurement by means of which the presence and degree of specific deficiencies may be noted. Five cases are described in connection with interpretation of a Diagnostic sheet. The two brief chapters in Part III are devoted to a discussion of the principles relating to remedial measures, followed by a bibliography of 20 numbers and an appendix on statistical methods.

L. Z.

---

5. *Growth Curves Obtained from an Objective Study of Reading Habits.*—Some of the limitations of standardized tests are aptly demonstrated by the type of measurement undertaken by Dr. Buswell[2] and reported in the second Chicago monograph over his signature.

---

[1] Gray, C. T.: "Deficiencies in Reading Ability." D. C. Heath and Co., Boston, 1922, pp. XIV + 419.

[2] Buswell, Guy T.: Fundamental Reading Habits: A Study of Their Development. *Supplementary Educational Monographs*, No. 21, University of Chicago, Chicago, 1922, pp. XIV + 150. $1.50.

The development of fundamental eye habits is traced through the various growth stages from earliest attempts to read in the first grade to the maturity exemplified by college seniors. In a detailed analysis of first grade reading records made at stated intervals by the use of a further improvement on the apparatus used by C. T. Gray and others, various approaches to reading are contrasted and compared psychologically.

In order to keep other factors constant all subjects above the first grade read the same material. For each of 179 cases the records show the average number of fixations per line, average duration of eye-pauses and average number of regressive movements per line. Growth curves show the development in each of the elements listed above. Individual variations resulting from various cases are an index of needs which specific training should supply. The author points out the danger of evaluating any method from a skill measure taken at an early growth stage and shows that devious paths of growth may lead to maturity, although some necessitate needless meanderings. This study should lead to further research and did in the scientific selection of procedure. The attempt to get objective indices of growth in *habits* and attitudes is encouraging to those who count such values at least as significant as skill.

L. Z.

---

6. *A Case Book in the Reading Field.*—Clinical Study as a method of education and pursued by several members of the Chicago group. Paralleling the technic of medical research and social investigation Dr. W. S. Gray[1] and his co-workers studied 27 individual cases to determine the significant characteristics of poor readers, causes of difficulty, and appropriate remedial instruction. A study of each child's history was made by compiling available school records, and consulting teachers and parents. In addition to an analysis of the results of standardized reading tests and the use of intelligence tests the specific nature of each child's difficulty was painstakingly determined by the aid of special tests, the use of laboratory equipment, analysis, observa-

[1] Gray, William S. with the cooperation of Delia Kibbe, Laura Lucas, Lawrence W. Miller: Remedial Cases in Reading: Their Diagnosis and Treatment. *Supplementary Educational Monographs*, No. 22, University of Chicago, Chicago, 1922, p. 208. $1.75.

tion and interview. Individual remedial instruction was organized under carefully controlled conditions. After approximately 2 months of prescribed training during which individual reactions were recorded, the pupils were re-tested.

The monograph reports before and after scores of each case together with a detailed record of procedures. Cases are grouped with reference to the type and cause of difficulty. Pupils who make little or no progress in reading are most easily detected but the causes for such deficiency were found hardest to determine because of the variety of possible contributing factors. Such pupils need a special type of instruction. It is highly desirable that they be recognized early, because their disability is often due to physical or intellectual limitations.

In the next group are included all pupils who encountered serious difficulty in interpretation. Although the causes for such deficiency were found to be numerous and the remedial procedure had to be varied to suit the conditions of each case, it is significant to note that effective instruction resulted in the removal of the cause.

Pupils who encountered difficulty in the mechanics of reading, form the third group. Ten separate causes were isolated and effective remedial measures were necessarily different and related to particular causes. Pupils whose rate of silent reading was unsatisfactory are next considered. Gray concludes that undue emphasis on rate before rudimentary habits are established may easily increase difficulties of recognition and prevent some pupils from becoming fluent readers.

Remaining cases are put together as weak in all phases of reading. In such cases individual remedial work is especially urgent and helpful if based on a recognition of the varied needs.

The monograph concludes with a brief chapter on similar studies carried on in a city school system and will no doubt be instrumental in bringing the possibility and value of such work to the attention of other progressive communities.

Educational practitioners will find Dr. Gray's case book suggestive in the diagnosis of thousands of other cases in which the "symptoms" are somewhat similar. Cumulative records of results add, determine, and reveal the validity of remedial measures and provide new data.